U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

PB-253 981

PROCEEDINGS NUMBER 2 OF THE ORD ADP WORKSHOP HELD AT THE EPA GULF BREEZE ENVIRONMENTAL RESEARCH LABORATORY, FLORIDA ON NOVEMBER 11-13, 1975

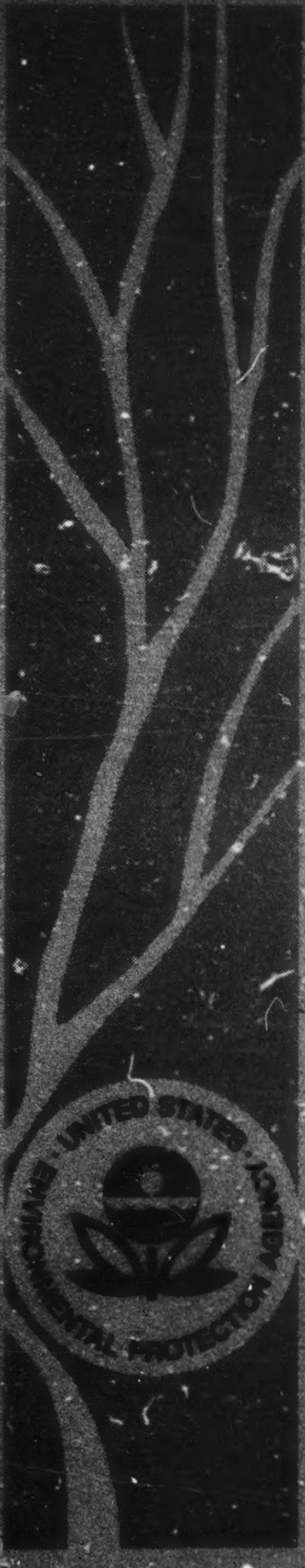
ENVIRONMENTAL PROTECTION AGENCY

APRIL 1976

177059

PB 253 981

PROCEEDINGS No. 2



ORD
ADP
WORKSHOP

November 11-13, 1975

Office of Research and Development
U.S. Environmental Protection Agency
Washington, D.C. 20460

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

This document is available from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Do not order from the U.S. Environmental Protection Agency.

DISCLAIMER

This report has been reviewed by the Office of Research and Development, U.S. Environmental Protection Agency, and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/9-76-008	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE ORD ADP WORKSHOP PROCEEDINGS NO. 2		5. REPORT DATE April 1976 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Denise Swink		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Office of Monitoring & Technical Support Office of Research and Development U.S. Environmental Protection Agency Washington, D.C. 20460		10. PROGRAM ELEMENT NO. 11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Same as above		13. TYPE OF REPORT AND PERIOD COVERED Symposium Proceedings, Nov. 11-13, 1975 14. SPONSORING AGENCY CODE EPA/ORD

15. SUPPLEMENTARY NOTES

The ORD ADP workshop was held Nov. 11-13, 1975 at EPA Gulf Breeze Environmental Research Laboratory, Gulf Breeze, Florida

16. ABSTRACT

The second ORD ADP workshop focused on the merits of past data acquisition and manipulation techniques and procedures, and provided suggestions for new approaches from the views of both providers and users of ADP resources. Subject area included instrumentation, process control, and data management aspects of laboratory automation; strengths and weaknesses of scientific analysis; utility of environmental data bases; developments in remote sensing projects, and future developments in EPA ADP resources.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Electronic Computers Laboratories Instruments Process Control	Laboratory Automation Data Management Scientific Analysis Environmental Data Bases Remote Sensing Techniques	09/B 14/B

PRICES SUBJECT TO CHANGE

18. DISTRIBUTION STATEMENT UNLIMITED	19. SECURITY CLASS *(This Report)* UNCLASSIFIED	21. NO. OF PAGES 205
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE 7.75-2.25

i.

EPA Form 2220-1 (9-73)

☆USGPO: 1976 — 657-695/5418 Region 5-11

EPA-600/9-76-008
April 1976

ORD ADP WORKSHOP PROCEEDINGS
NO. 2

Sponsored By

Denise Swink, ORD ADP Coordinator
Technical Information Division (RD-680)
Office of Monitoring and Technical Support
Office of Research and Development

U.S. ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF RESEARCH AND DEVELOPMENT
WASHINGTON, D.C. 20460

ia

FOREWORD

The second ORD ADP Workshop was held November 11-13, 1975, at the EPA Gulf Breeze Environmental Research Laboratory, Gulf Breeze, Florida. This workshop focused on the merits of past data acquisition and manipulation techniques and procedures, and provided suggestions for new approaches from the views of both providers and users of ADP resources. Participating in the workshop, sponsored by the Office of Research and Development, were representatives of EPA Headquarters program offices, Regional offices, laboratories, and organizations outside of EPA.

Denise Swink
ORD ADP Coordinator

Preceding page blank

TABLE OF CONTENTS

Page Number

FOREWORD iii

AGENDA 1

Opening Remarks
Denise Swink 5

Welcome Address
Tudor T. Davies 6

Keynote Address
Wilson K. Talley 7

Microprocessors
D.M. Cline 8

A Flexible Laboratory Automation System for an EPA Monitoring Laboratory
Bruce P. Almich 10

Data Acquisition System for an Atomic Absorption Spectrophotometer, an Electronic Balance, and an Optical Emission Spectrometer
Van A. Wheeler 19

An Automated Analysis and Data Acquisition System
Michael D. Mullin 25

A Turnkey System: the Relative Advantages and Disadvantages
D. Craig Shew 31

One Approach to Laboratory Automation
Jack W. Frazer 33

Software Compatibility in Minicomputer Systems
John O.B. Greaves 38

Summary of Discussion Period - Panel I 40

Laboratory Data Management
William L. Budde 42

Suspended Particulate Filter Bank and Sample Tracking System
Thomas C. Lawless 44

Preceding page blank

Page Number

Eight Years of Experience With an Off-Line Laboratory Data Management System Using the CDC 3300 at Oregon State University
D. Krawczyk 50

The CLEANS/CLEVER Automated Clinical Laboratory Project and Data Management Issues
Sam D. Bryan 60

Requirements for the Region V Central Regional Laboratory (CRL) Data Management System
Billy Fairless 65

Data Collection Automation and Laboratory Data Management for the EPA Central Regional Laboratory
Robert A. Dell, Jr. 74

Sample Management Programs for the Laboratory Automation Minicomputer
Henry S. Ames and George W. Barton, Jr. 76

Summary of Discussion Period - Panel II 79

The State of Data Analysis Software in the Environmental Protection Agency
Gene R. Lowrimore 81

National Computer Center (NCC) Scientific Software Support - Past, Present, and Future
M. Johnson 84

Exploitation of EPA's ADP Resources: Optimal or Minimal?
John J. Hart 86

Scientist, Biometrician, ADP Interface
Neal Goldberg 89

Statistical Differences Between Retrospective and Prospective Studies
Dr. R.R. Kinnison 91

Raising the Statistical Analysis Level of Environmental Monitoring Data
Wayne R. Ott 93

Quality Assurance for ADP (and Scientific Interaction With ADP)
R.C. Rhodes 95

How to Write Better Computer Programs
Andrea T. Kelsey 98

Summary of Discussion Period - Panel III 103

The Utility of Bibliographic Information Retrieval Systems
Johnny E. Knight 106

Biological Data Handling System (BIO-STORET)
Cornelius I. Weber 109

Page Number

Utility of STORET
C.S. Conger 111

The Uses and Users of AEROS
James R. Hammerle 114

Description and Current Status of the Strategic Environmental Assessment System (SEAS)
C. Lawrence 118

Improving the Utility of Environmental Systems
Donald Worley 121

Summary of Discussion Period - Panel IV 123

Operational Characteristics of the CHAMP Data System
Marvin B. Hertz 125

Development of Thermal Contour Mapping
George C. Allison 135

Remote Sensing Projects in the Regional Air Pollution Study
R. Jurgens 138

Automatic Data Processing Requirements in Remote Monitoring
J. Koutsandreas 148

Developments in Remote Sensing Projects
Sidney L. Whitley 156

Summary of Discussion Period - Panel V 173

Agency Needs and Federal Policy
Melvin L. Myers 175

Minicomputers: Changing Technology and Impact on Organization and Planning
Edward J. Nime 178

Univac 1110 Upgrade
M. Steinacher 181

A Case for Midicomputer-Based Computer Facilities
D. Cline 183

Status of the Interim Data Center
K. Byram 185

Large Systems Versus Small Systems
R.W. Andrew 187

	Page Number
Summary of Discussion Period - Panel VI	189
APPENDIX - List of Attendees	A-1

AGENDA

ORD ADP WORKSHOP NO. 2
NOVEMBER 11-13, 1975
GULF BREEZE, FLORIDA

Opening Remarks	Denise Swink
Welcome Address	T. Davies
Keynote Address	W. Talley

Panel I

Laboratory Automation - Instrumentation and Process Control

Chairman: D. Cline

Microprocessors	D. Cline
A Flexible Laboratory Automation System for an EPA Monitoring Laboratory	Bruce P. Almich
Data Acquisition System for an Atomic Spectrophotometer, an Electronic Balance, and an Optical Emission Spectrometer	Van A. Wheeler
An Automated Analysis and Data Acquisition System	Michael D. Mullin
A Turnkey System: The Relative Advantages and Disadvantages	D. Craig Shew
One Approach to Laboratory Automation	Jack W. Frazer
Software Compatibility in Minicomputer Systems	John Greaves
Question and Answer Period – Panel I	

Panel II

Laboratory Automation - Data Management

Chairman: William L. Budde

Laboratory Data Management	William L. Budde
Suspended Particulate Filter Bank and Sample Tracking System	Thomas C. Lawless
Eight Years of Experience With an Off-Line Laboratory Data Management System	D. Krawczyk
The CLEANS/CLEVER Automated Clinical Laboratory Project and Data Management Issues	Sam Bryan

Requirements for the Region V Central Regional Laboratory (CRL) Data Management System	Billy Fairless
Data Collection Automation and Laboratory Data Management for the EPA Central Regional Laboratory	Robert A. Dell, Jr.
Sample Management Programs for the Laboratory Automation Minicomputer	George Barton
Question and Answer Period – Panel II	

Panel III

Strengths and Weaknesses of Analysis of Scientific Data in EPA

Chairman: G. Lowrimore

The State of Data Analysis Software in the Environmental Protection Agency	Gene R. Lowrimore
National Computer Center (NCC) Scientific Software Support - Past, Present, and Future	M. Johnson
Exploitation of EPA's ADP Resources: Optimal or Minimal?	John J. Hart
Scientist, Biometrician, ADP Interface	Neal Goldberg
Statistical Differences Between Retrospective and Prospective Studies	R.R. Kinnison
Raising the Statistical Analysis Level of Environmental Monitoring Data	Wayne R. Ott
Quality Assurance for ADP (and Scientific Interaction With ADP)	R.C. Rhodes
How to Write Better Computer Programs	Andrea Kelsey
Question and Answer Period – Panel III	

Panel IV

Utility of Environmental Data Systems

Chairman: Johnny E. Knight

The Utility of Bibliographic Information Retrieval Systems	Johnny E. Knight
Biological Data Handling System (BIO-STORET)	Cornelius I. Weber
Utility of STORET	C.S. Conger
The Uses and Users of AEROS	James R. Hammerle
Description and Current Status of the Strategic Environmental Assessment System (SEAS)	C. Lawrence
Improving the Utility of Environmental Systems	Donald Worley

Non-Use of EPA Data Systems	D. White
Question and Answer Period – Panel IV	

Panel V

Developments in Remote Sensing Projects

Chairman: Marvin B. Hertz

Operational Characteristics of the CHAMP Data System	Marvin B. Hertz
Development of Thermal Contour Mapping	George C. Allison
Remote Sensing Projects in the Regional Air Pollution Study	R. Jurgens
Automatic Data Processing Requirements in Remote Monitoring	J. Koutsandreas
Developments in Remote Sensing Projects	Sidney L. Whitley
Question and Answer Period – Panel V	

Panel VI

Future Developments in ADP Resources for EPA

Chairman: Melvin L. Myers

Agency Needs and Federal Policy	Melvin L. Myers
Minicomputers: Changing Technology and Impact on Organization and Planning	E. Nime
Univac 1110 Upgrade	M. Steinacher
A Case for Midicomputer-Based Computer Facilities	D. Cline
Status of the Interim Data Center	K. Byram
Large Systems Versus Small Systems	R.W. Andrew
Question and Answer Period – Panel VI	

OPENING REMARKS

By Denise Swink

I am pleased to see the familiar as well as the new faces present here today. Such representation at this meeting substantiates the need for, and usefulness of, the ORD ADP workshop series. To refresh memories and provide background for those of you who did not participate in the first workshop held in October 1974, I will summarize the purpose, subjects covered, and impacts of the first workshop.

It became apparent in 1974 after my first year of functioning as the ORD ADP Coordinator that the scientific community of EPA involved in data processing applications had few mechanisms to transfer or find significant ADP technology. In response to the need for better communication, the first workshop was designed to promote state-of-the-art techniques as well as the sharing of experience and knowledge in the area of scientific applications and processing of data. The subjects presented in formal papers included: mathematical, scientific, and statistical applications software; applications of minicomputers; applications of interactive graphics; laboratory data management; chemical information systems; capabilities of the Univac 1110, and ADP policies. Participation at this workshop included not only personnel from the Office of Research and Development but also Regional offices, Headquarters program offices, and organizations outside of EPA. The *Proceedings of the ORD ADP Workshop No. 1* are available from the National Technical Information Service (NTIS) and can be purchased on request using the accession number PB241 150/AS.

The first workshop was successful; however, the participants commented that they thought it would be beneficial as a follow-on activity to spend time addressing issues and operational problems associated with the subjects covered. Hence, the second workshop has been designed to respond to this gap of information transfer.

Since ADP activities rely on an interdependent network of providers and users, many problems and questions arise concerning methods and policies. To manage available ADP resources effectively, one must discern the level of technology and resources appropriate for an application. Because of constantly accelerating developments in technology coupled with the provision of ADP resources from several organizations with differing management philosophies, it becomes extremely difficult to optimize one's use of ADP resources. Consequently, this workshop will focus on the merits of past approaches and provide suggestions for new approaches from the view of both a provider and a user.

To maximize participation, this workshop is organized into six panel sessions for which there will be short presentations by each panel member with a question and answer period immediately following the presentations. The panels have been established to cover the topics of: instrumentation and process control aspects of laboratory automation, data management aspects of laboratory automation, strengths and weaknesses of scientific analysis of data in EPA, utility of environmental data bases, developments in remote sensing projects. and future developments in ADP resources for EPA. *Proceedings* from this workshop also will be available from NTIS by April 1976.

In closing, I thank you for your interest and participation in the ORD ADP workshops and look forward to our increased and improved communications in the future.

Preceding page blank

WELCOME ADDRESS

By Tudor T. Davies

I am delighted to welcome you to the ADP Conference and to the Pensacola area, particularly to the Gulf Breeze Environmental Research Laboratory. Besides the obvious benefits of being in this equable climate, we are fortunate that many people recognize this and use it as a site for Agency meetings. Many of you are old friends, but there are some present who will be unfamiliar with the capabilities and the research program of the Gulf Breeze Laboratory. Therefore, we would like to invite you to visit us tomorrow.

When I was a newcomer to the Agency and becoming acquainted with the Great Lakes community of research, Regional, and State people, one of the common factors that tied the water people together was the STORET system. Many voices have been raised against it. I believe, however, that these people are speaking from ignorance about its capabilities. I strongly support it as the best available interactive data storage system. With the evolution of the BIOSTORET system, I feel that we have an excellent comprehensive data system which is very much user-oriented. We should strongly defend and support its continuance and future growth. Although many of the topics to be addressed in this workshop reflect our interest in automating laboratory analysis and controlling sample flow and data arrangement, the eventual use of the data and its communication are perhaps most significant to us in an overall sense. At Grosse Ile, we found that data storage was an expensive business but, once stored in the STORET system, it was available for sample analysis and complex modeling exercises by the whole user community. The community action in EPA toward ADP is most encouraging.

KEYNOTE ADDRESS

By Wilson K. Talley

The rhetoric used when EPA was formed is still valid, that is, pollution problems are too often perceived in isolation and addressed in isolation. The result is the suboptimization of our society's dealing with the environment. Yet, EPA was structured to be a regulatory agency that was required to take a total view.

In the early days, the first half of the 1970's, the Agency attacked the most obviously serious, and least controversial, problems. Pollutants and classes of pollutants were identified and abated. Today, we have exhausted the single pollutant/single medium/single species approach. To continue to enhance and protect environmental quality, we must consider the residuals of our activities.

This current situation presents a challenging opportunity for us to accept. For example, within the last year, we have imposed on ourselves the requirement that we submit Environmental Impact Statements with respect to our Agency decisions. Let us contrast our Agency's view with the narrower missions discharged by other Federal agencies. The Department of Agriculture must maximize the production of food and fiber. One way for discharging its mission would be to assume that there is an abundance of inputs, such as the land itself, capital, labor, chemicals, energy, and water, and that the residuals are unimportant. Only when an input runs short or a residual becomes a problem is it necessary to consider the social, economic, or environmental systems outside agriculture.

Another example might be the Department of Health, Education, and Welfare's important mandate for a health delivery system. Originally that system was viewed primarily as a health services delivery system. That view is tenable only as long as the social cost of a solution is not out of proportion when based on taking the present system and making it bigger.

We recognize that EPA cannot do its job and those of its sister agencies. But note that EPA can do its job more easily and completely if it works with those agencies. And in doing so, there is every reason to believe that the other agencies will be able to do a better job with their primary missions. For instance, a large portion of the problem with agriculture chemicals is their misuse. One investigator has estimated that only 20 percent of the average pesticide application is effective. Another points out that for some crops on some lands, fertilizers have passed the point of diminishing returns. Changes in irrigation practices could not only save water initially but could also cut down on the runoff pollution by salts. With respect to a health delivery system, a more correct approach is to regard environmental protection as a necessary adjunct to health maintenance.

Similar examples abound that indicate a return to the old conservation ethic: use less, use it well, and you will waste (pollute) less. Our concern should include not only present missions but also future problems. We should anticipate and move against these future problems to eliminate them at the least cost. However, without the proper data base or the tools to manipulate it into information, our task will be impossible.

ADP may provide us one of the tools we need to realize the full potential of the Agency. This workshop is a continuing effort to provide us with appropriate tools. It is being held but three weeks shy of EPA's fifth birthday. The Agency is using that anniversary as an opportunity to review the first five years and to plan for the future. This workshop can make a valuable contribution to that future.

MICROPROCESSORS

By D. M. Cline

INTRODUCTION

In recent years, system logic design engineers have included minicomputers as integral parts of various process control and analytical instrumentation applications. The relatively high cost of minicomputers has inhibited their widespread use in these types of applications. As an alternative, system designers have used "random logic" techniques in which the expense of a single system is reduced by volume production. Then in the early 1970's, a phenomenal new electronic device, the microprocessor, was produced as a result of large scale integration (LSI) technology. The microprocessor unit is a central processing unit on a chip of plastic with typical dimensions of 0.25 by 0.25 inches. It includes the functional units of arithmetic logic, control circuitry, and registers. The unit alone does not constitute a microcomputer, since it requires two additional components: input/output (I/O) circuitry and memory. Although the microprocessor was initially used by the system logic designer, new and improved microprocessor designs are beginning to be utilized by users other than engineers. Regardless of profession, this new technology is demanding that the system developer have expertise in both hardware and software techniques.

CHARACTERISTICS

Word size and speed are the first characteristics about which a potential microprocessor user inquires. However, if the interrogation ends at this point, the user will have many surprises in store as system development proceeds. For instance, the microprocessor may not include interrupt circuitry, direct memory access (DMA) ability, accessible stack, or BCD arithmetic capability. The component that provides system timing, the clock, initially was composed of circuitry external to the microprocessor unit, but some of the more recently introduced microprocessors include clocks as integral portions of the chips themselves. Many microprocessors do not have clocks and their manufacturers do not offer them on a separate chip; therefore, the user has to design and construct a single- or two-phase clock as required for operation of the microprocessor.

Many of the microprocessors are supplied by a single manufacturer. This is an important factor to consider when system life may be long and one must be assured of spare or replacement components. Another factor to consider is the number of power supply voltages required. Most models require two, but some newer models require a single positive 5-volt supply and are compatible with the TTL family of logic. One should also consider both the number and the kind of registers available.

In the past year, several manufacturers and independent suppliers have begun offering microprocessor kits which include the microprocessor and the required support logic. This could be useful for prototype development because many include software and logic for a serial device. The three major sources of kits are the microprocessor manufacturers, the electronics distributors, and the system houses. The system houses seem to offer the most complete kits, which include the microprocessing unit, support logic, printed circuit boards, power supplies, cabinet, programer's console, switches, and lights.

One major disadvantage of programing a microprocessor is that the peripheral devices required for expedient software development are more expensive than the microprocessor itself. Thus, it is very difficult to produce a single system at a low cost which utilizes a microprocessor even though the components which comprise the system are inexpensive. One last characteristic, which manufacturers of microprocessors are finding to be of considerable marketing value, is a chip that has an instruction set compatible to a popular minicomputer.

SUPPORT LOGIC

Support logic consists of many devices, including memories (ROM), random access memories (RAM), programable read-only memories (PROM), electrically programable read-only memories (EPROM), clocks, shift registers, and parallel and serial I/O interfaces. Most microcomputer systems contain at least one MPU, one ROM, and one RAM. The ROM is a device from which information can be read but on which information can not be written. Usually the control logic or "computer program" is implemented in ROM because its contents are not lost when power is removed. Since a RAM is a read/write memory device, it is used for data storage and its contents are lost when power is removed. The PROM is a device that can be irreversibly programed under special conditions but acts like a ROM once it is programed. The EPROM is similar to the PROM but can be reprogramed under special conditions.

APPLICATIONS

Microprocessor applications abound. Microprocessors are used as traffic light controllers, electric range controllers, numerical controllers, and elevator controllers. They are used as the control logic for slower computer peripherals such as cassette drives, flexible disk drives, line printers, card readers, and plotters. They are used in point-of-sales terminals, cash registers, adding machines, and fare collection devices.

A recent article in *Computer World* reported that IMS Associates, Inc. had combined multiple Intel 8080 microprocessors into an array configuration, the smallest containing 32 Intel 8080 MPUs and the largest containing 512 Intel 8080 MPUs.[1] The systems are reported to offer high computing power at a low cost.

In this paper, a brief overview of a new and exciting technology has been presented, including some of the characteristics of microprocessors as well as some of the pitfalls to avoid when selecting a microprocessor. Microprocessor-based systems provide the system designer with the opportunity to reduce costs, component count, and size. However, an essential knowledge of the hardware and software characteristics of the microprocessor is necessary in order to utilize the microprocessor effectively in a system.

Reference

1 Frank, Ronald A., "IMSAI Arrays Micros for Low-Cost Power," *Computer World*, October 1975.

A FLEXIBLE LABORATORY AUTOMATION SYSTEM FOR AN EPA MONITORING LABORATORY

By Bruce P. Almich

INTRODUCTION

In the environmental field the measurement of specific air and water pollutant chemicals is a very important activity.[1] Until reliable measurements are made and correlated with undesirable health or wildlife population effects, environmental concerns are limited only to those concerned with purely aesthetic values. Currently, there is considerable emphasis on setting standards for acceptable air and water quality, issuing permits for discharge of wastes into rivers and oceans, monitoring these effluents to ensure compliance with permit limitations, and conducting enforcement actions when violations occur. All of these activities are increasing the demand for more and improved chemical environmental analyses.

Improved analyses embody accuracy and precision, and require extensive use of analytical quality control techniques.[1] Quality control is often omitted in analytical laboratories because of its cost and time requirement. With this omission, the meaningfulness of the measurements decreases substantially. There is nothing more costly than the wrong answer. Another aspect of better analysis is the desire for new kinds of measurements that are more revealing about the state of environmental pollution than that provided by traditional measurements. These more revealing measurements are often more complex and simply cannot be accomplished economically, or at all, without some form of automation.

With the above remarks in mind, one can summarize the objectives of laboratory automation for EPA as follows:

- Increase instrument and laboratory throughput for a given level of instrumentation and operations personnel
- Increase the productivity of laboratory personnel
- Improve accuracy and precision of analytical results with instream data quality control and instrument reliability assurance procedures
- Reduce clerical time and errors by eliminating manual calculations and transcriptions of data
- Reduce tedium by automating as many repetitious laboratory tasks as practicable
- Incorporate instruments and techniques into laboratory operations which would not otherwise be practical or possible due to technical or economic constraints
- With all costs considered, provide a substantial positive net benefit to the laboratory for the lifetime of the added equipment.

Many of these objectives are fully discussed elsewhere.[2 3]

In response to the needs and objectives stated here, the Environmental Monitoring and Support Laboratory (ORD-EMSL) and the Computer Services and Systems Division (OPM-CSSD) of the Cincinnati EPA laboratories have been conducting a project in concert with the Lawrence Livermore Laboratory (ERDA-LLL). The result has been the development of a highly flexible laboratory automation system capable of meeting the above objectives for a wide variety of environmental monitoring laboratories. EMSL was chosen as the site for pilot system installation and integration because its primary mission is: "To develop, improve, and validate methodology for the collection of physical, chemical, radiological, microbiological, and biological water quality data by EPA Regional offices, Office of Enforcement and General Counsel, Office of Air and Water Programs, and other EPA organizations."[1] In the direct support and initial direction of this project, CSSD has been carrying out its major responsibilities: "Provide EPA Cincinnati focal point for coordination and integration of computer systems across technical lines.... plan, coordinate, and carry out a program for exploitation of scientific and technical application of computers to EPA needs."[4] LLL was chosen to assist in the project because of its previous record of solid accomplishments in relevant areas as well as its existing highly qualified technical and professional staff.

PROJECT GOALS

In order to meet the objectives and needs of EPA monitoring laboratories, a thorough systems analysis approach was taken to establish the exact goals of the project, to write detailed specifications for hardware and software, and to develop an implementation plan.[2 3] Several of the goals defined as mandatory include:

- To develop a laboratory automation system that would incorporate presently owned chemical analysis instrumentation widely used throughout the Agency for measuring water quality parameters.

- To develop this methodology to permit the adaptation of the technology to other EPA laboratories at very significant cost and time savings. In particular, designs for hardware interfaces between instruments and computers, as well as all custom computer software, would become public property to be used in any EPA laboratory without further development or licensing costs.

- To develop an open-end design for both hardware and software that would permit the attachment of many additional instrument types for measurements, including nonwater parameters. This goal includes the minimum-cost ability to have varying numbers of each automated instrument type for a given laboratory, depending only on laboratory needs.

- To take advantage of presently available computation power in writing as much of the software as possible in a very flexible, high-level, modern programing language. This would assist scientific personnel in modifying and improving software, and facilitate the transfer of technology to other laboratories at minimum cost.

- To design the system with sufficient flexibility so that it is applicable to methods development research as well as to the production atmosphere. In an automated environment, careful testing of new procedures becomes economically and technically possible with a statistically significant number of samples.

- To maximize flexibility but minimize redundancy and inefficiency, from the viewpoint of an Agency-wide computer software effort.

It was recognized early in this project that a satisfactory level of technical and administrative coordination and communication would be necessary for EPA to realize the goals and objectives of laboratory automation. To this end, the following relationship attributes were defined for LLL, CSSD/EMSL, and other "client" laboratories:

- Design efforts for major system components and facilities would proceed as an "interactive, iterative process" among the interested parties.[3]

- The first stage of this process would involve a thorough system specification and design resulting from the collection of requirements from interested parties throughout EPA's monitoring laboratory community.

- LLL would assume technical leadership in the initial systems level hardware and software implementations, producing sufficient documentation to allow EPA to independently maintain and modify the delivered turnkey systems at all technical levels.

- Following the successful installation and debugging of three turnkey systems, EPA would take a more active technical role in the areas of maintenance, new equipment additions, and subsequent client laboratory implementations. With the continued assistance of outside sources such as LLL, EPA would use its increased level of in-house expertise and coordination to maintain existing systems and to add to them, whenever applicable.

- At maturity, the effort would require a level of continued EPA technical and administrative interaction to the point that the formation of an Agency-wide users' group would be indicated.

PROJECT STATUS

Instrumentation

Since the last report of the project status to this workshop, two complete hardware installations of these systems have become operational, with a third due on March 1, 1976.[1] The number and types of instruments completed, together with their performance characteristics and controlling software, are substantially identical to the original intentions of the design.[1,2] The major exceptions to this are the following:

- A furnace has been added by LLL to the Automated Atomic Absorption spectrometer system at the CRL-V Chicago site.
- The hardware and controlling software for a Mettler Balance system at CRL-V has been completed by LLL.
- EMSL-Cincinnati is completing the addition of another "new" instrument to the system; i.e., a Coleman 124 Spectrophotometer.
- With a CRL-V implementation installed, several variations of automatic sample changers are now available for the systems. Capacities start at 40 samples.
- CSSD-Cincinnati has responded to CRL-V in their priority requirement for a remote job entry capability, allowing the laboratory computer to send and receive batch mode jobs in conjunction with EPA's IBM and Univac computing centers. The technical aspects of this addition are complete.

For convenience, the existing and planned hardware for this project has been broken down into a number of functional areas and is presented in Table 1.

Software

Table 2 illustrates the various types of software included and planned for each laboratory automation system installed as part of this project. The present status of the software includes performance meeting the original specifications,[1,2] with the notable exception that the Sample File Control systems and applications programs are still in the design phase. Further information on software status will be available after February 1, 1976, from the author. In brief, however, the presently running software includes all data acquisition, reduction, instrument control, and quality assurance procedures that were originally intended for the existing automated instruments. In addition, a variety of other BASIC applications programs are either presently available or are under test/debug phases at the various laboratory sites. Present capabilities include the production of printed laboratory analysis reports suitable for filling and reporting in the usual manner.

In order to assist the reader in visualizing the "total" software picture for this system, a core map for the EMSL-Cincinnati site is given in Figure 1. A hardware memory protection unit functionally divides the available memory into three partitions: foreground, background, and operating system area. Although one can subdivide the available 64k of address space in many ways, it was found that the allocations shown were the most optimal for the purposes of most laboratory systems, with the accompanying constraint that no partition can be larger than 32k. Thus, the foreground partition contains Multiuser Extended Basic, the LLL assembly language instrument drivers, and the user data area shared by each user via the swapping disk. The background partition is intended for the operating portions of the Sample File Control programs as well as an area for utility functions and a limited amount of program development. The operating system, MRDOS revision level 3.02, resides in the lowest 16k or so of core space. For the given revision levels of the DGC-supplied software shown, the core map represents not only the optimal configuration but also the maximum total amount of core that should be used for the system.

As one shifts from the design/development of this project to maintenance, it is necessary to reevaluate the validity of having a sophisticated operating system such as MRDOS resident in each laboratory automation computer system. With the advantages and disadvantages cited in Table 3, it is clear that advantages are apparent when the system is under development and testing. However, the disadvantages tend to appear in the longer term as maintenance costs for system level software upgrades. This type of upgrade is required in order to maintain pace with software and hardware technology as well as to keep the computer vendor's hardware and software support for the "current" revision level. The degree to which the technical and economic aspects of the disadvantages can be minimized varies with the number of "custom" systems level software interfaces that must be adjusted with each revision level of the vendor-supplied code. This is one reason why the LLL designers have maintained a "hands-off" policy with respect to custom modifications of the MRDOS software. There are, however, a fair number of other aspects of the total software picture which are showing sensitivity to revision levels as time passes. The costs of maintaining the entire software system, therefore, can be minimized only if these software modules are maintained as close to identical as possible for all installations of these systems. Thus, changes can be made in sensitive areas once, and only once, with timely distribution throughout the Agency.

Table 1
Summary of Existing and Planned Hardware Types*

A. Instruments

1. Atomic absorption spectrometers, furnace options
    - o P.E. 503, 303, 306
    - o I.L. 453, Varian AA-5
2. Beckman total organic carbon analyzer
3. Technicon autoanalyzers
    - o Single/multiple channel
    - o Types I and II
4. Jarrell-Ash 3.4 meter electronic readout emission spectrometer
5. Varous automatic sample changers, capacity of 40 and up
6. Mettler balance
7. P.E. (Coleman) 124 double beam spectrophotometer
8. Various instruments similar to the above*
9. High data volume instruments already fitted with digital control systems (e.g., GC-mass spectrometer, plasma emission spectrometer*)
10. Various instruments characterized by low cost, infrequent data production, etc. (e.g., pH meter, turbidmeter, etc.*)

B. Computer-Related

1. CPU: Data General Nova 840 (or equivalent instruction set emulation capability at increased performance*)
2. Memory: 64k words (up to 128k words*) of high-speed core (or interleaved semiconductor*)
3. Arithmetic: Hardware multiply/divide and floating point processor
4. Peripherals
    - o Fixed head swapping disk
    - o Removable moving head data and program storage disk
    - o Tape drive for backup and data archives
    - o Digital interface and up to 32 channel analog to digital converter
    - o Medium-speed line printer
    - o Low-speed hard copy and CRT user and computer control terminals
    - o Custom-fabricated instrument interfaces
    - o Asynchronous telecommunications interface and modems
    - o Synchronous telecommunications interfaces*
    - o Dual processor - shared disk adapters*

*Planned hardware types are indicated with asterisks.

Table 2
Summary of Existing and Planned Software Types*

A. Data General Supplied, DGC/EPA Maintained

1. Operating system: MRDOS revision 3.02 (4.02, 5.xx*)
2. Data acquisition and control: Multiuser Extended BASIC (up to 16 users), revision 3.6 (4.xx, up to 32 users*)
3. Utilities: assemblers, loaders, debuggers, editors, command language, Fortran IV, V, and Remote Job Entry package (RJE-IBM)

B. LLL Supplied, LLL/EPA Maintained

1. Real-time assembly language routines and interface to BASIC for instrument control and data acquisition
2. Patches to DGC BASIC
3. Initial BASIC applications program package, including instrument controllers, data acquisition, and quality control programs
4. System performance analysis and foreground/background communications packages

C. "Source-X" Supplied, EPA Maintained

1. Univac 1110 Remote Job Entry package (licensed from Gamma Tech.)
2. Sample File Control systems software package*
3. Sample File Control initial applications package*
4. Documented EPA internal software with broad application
5. All other software selected for EPA-wide support

D. Client Laboratory Supplied, Client Laboratory Maintained

1. Additions and changes to either BASIC or Sample File Control applications packages for incorporation of "local" needs
2. All other uncoordinated "local" changes to types A-C above* (case-by-case tradeoff with item C-5 above)
3. All other "new" local programs

Since the software involved in this maintenance function is deeply buried at the systems level, those who are knowledgeable about the subject believe that this will not impact on the flexibility of the systems. One can, therefore, conclude that the advantages of MRDOS in the laboratory clearly outweigh the disadvantages, even during the maintenance phases, if every effort is made to maintain Agency-wide compatibility for software items A-1 through A-3, B-1, B-2, and C-2 shown in Table 2.

Performance

To date, the hardware aspects of system performance have been quite good. Although insufficient operational experience has been accumulated at this point to

Table 3

Vendor-Supplied Real-Time Operating System Software in Laboratory Automation Computers

Advantages

- Efficient allocation of system resources:
  - Core: overlays, tasking, swapping, partitioning, reentrancy
  - CPU: dynamic scheduling, tasking, time slicing
  - Peripherals: file structuring, interrupt service, space management
  - Man-machine interface: console system control language
- Time-proven, reliable, vendor-supplied code
- Standardized utilities: compilers, editors, job stream managers, etc.
- Data and programs generally transportable among installations
- Compatibility with new and existing vendor-supplied hardware
- Program and system development accomplished at minimum cost

Disadvantages

- Operating system overhead can be significant: CPU time, core, cost
- Peripherals not directly accessible: 50 µsec overhead per interrupt
- User device driver implementation can become quite involved
- User responsibility to keep up with vendor-supplied upgrades of software; i.e., vendor discontinues support of outdated software releases
- Vendor software upgrades may require significant user software modifications and/or rewrites

recommend design changes in subsequent implementations, the following are included among present observations:

- The "problem child" for hardware maintenance has been the fixed head swapping disk. In two of the operating installations, this device has been quite unreliable. It is also the current performance bottleneck in the system because its low data transfer rate causes poor user response time during peak loads. Efforts are being made to improve this aspect of the system.

- One can take advantage of recent advances in computer technology in subsequent designs to include a cpu which is instruction-set compatible with the Nova 840 but also much faster and somewhat cheaper. The benefits include faster user response time in the face of the large number of automated instruments and program functions at the larger laboratories.

- The addition of new instrument types should proceed in an orderly manner so as not to impact the performance, reliability, or cost effectiveness of the existing system in an undue manner.

The software aspects of system performance also have been very good. Generally, it is believed that the conservative policy towards modification of systems level code has been responsible.[5] A second significant factor has been the level of creative thought brought to bear by LLL in the design and implementation of the custom software and hardware for the system. Nevertheless, obsolescence and the promise of improved response time for large numbers of operating instruments are presently encouraging an ongoing effort to upgrade the revision level of the systems software on an Agency-wide basis. Once the upgrade from revision 3 to 4 of MRDOS is performed during the first half of 1976, the following new features will be present:

- A new maximum core capacity of at least 128k words will be possible. The present 64k maximum may be a bit tight for the larger laboratories. Figure 2 shows an improved core map.

- A system tuning feature will be available for determining the size of the operating system partition on the basis of measured system performance for each laboratory site. This will allow for the true optimization of computer resources for each site, especially the larger installations, where the present state of optimization is pretty much a guessing game.

- Better features for spooling and BASIC system generation should improve system performance substantially.

PROBLEMS

As the non-Data Base Management aspects of this project approach maturity, the typical problems associated with technological innovation in the presence of budgetary and organizational constraints are evident. The basic issues include the following:

1. What is "standardization"? Management discusses it in broad, nontechnical terms, while people with technical responsibilities are grappling with the basic tradeoffs between "reinventing the wheel" and losing flexibility. A firm policy needs to be developed and accepted.

2. A formal process for adding new instruments and capabilities must be developed. Those who are close to a problem, tend to misestimate the broader implications and impacts of its solution. Those removed from a problem have difficulty fittin [illegible] into a priority scheme and must be motivated in obtaining a realistic, timely solution by those closer to the problem.

3. A mechanism for solving systems-level problems for vendor and LLL-supplies software is needed. Problems referred to DGC and LLL from a single point within EPA will be solved once and for all in a timely manner.

4. Progress towards a users' group should be initiated as soon as possible.

SUMMARY

A flexible laboratory automation system has been designed and currently is being proven as operationally viable within EPA. With the completion of three hardware installations and the first cut of sample file control programs during 1976, the previously stated goals and objectives will be satisfied to a large extent. The continued development of in-house Agency-wide expertise and experience in this field will facilitate not only the effective communication among the interested parties but also the increased capability to apply automation technologies to the solutions of laboratory problems.

REFERENCES

1 Budde, W.L., Nime, E.J., and Teuschler, J., "An Online Real-Time Multi-User Laboratory Automation System," *Proceedings No. 1, ORD ADP Workshop*, 1974.

2 Frazer, J. W. and Barton, G. W., "A Feasibility Study and Functional Design for the Computerized Automation of the Central Regional Laboratory EPA Region V, Chicago," *ASTM Special Technical Publication 578*, ASTM, 1975.

3 Frazer, J.W., "Concept of a Different Approach to Laboratory Automation," *Proceedings No. 2, ORD ADP Workshop*, 1975.

4 Nime, E. J., *CSSD Functional Responsibilities*, EPA Cincinnati, 1975.

5 Bunker, E., "If something works, don't fix it." "All in the Family" (TV series) 1974.

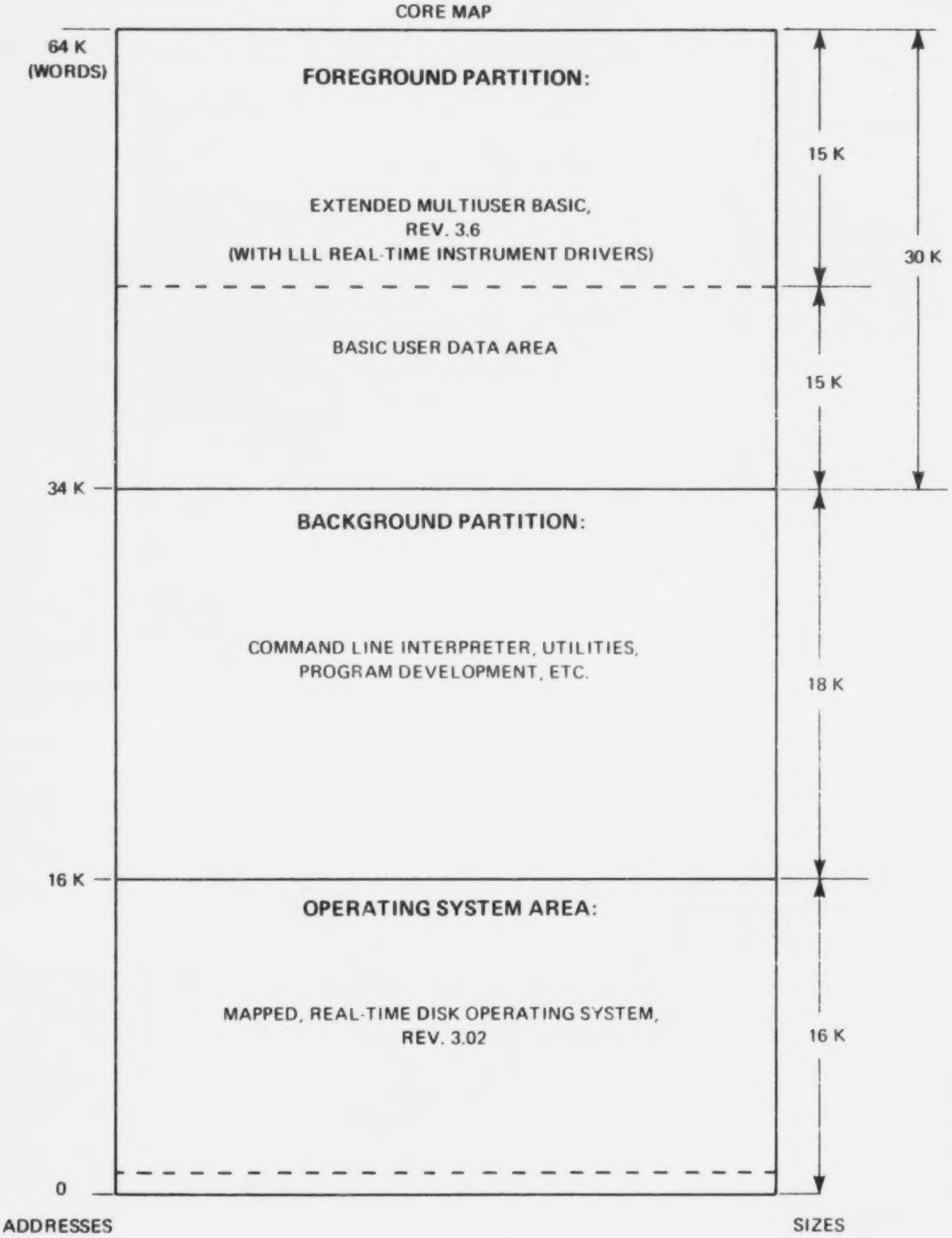

CORE MAP
64 K
(WORDS)
FOREGROUND PARTITION:
EXTENDED MULTIUSER BASIC,
REV. 3.6
(WITH LLL REAL-TIME INSTRUMENT DRIVERS)
15 K
30 K
BASIC USER DATA AREA
15 K
34 K
BACKGROUND PARTITION:
COMMAND LINE INTERPRETER, UTILITIES,
PROGRAM DEVELOPMENT, ETC.
18 K
16 K
OPERATING SYSTEM AREA:
MAPPED, REAL-TIME DISK OPERATING SYSTEM,
REV. 3.02
16 K
0
ADDRESSES
SIZES

Figure 1
Present Pilot Laboratory Automation System Configuration

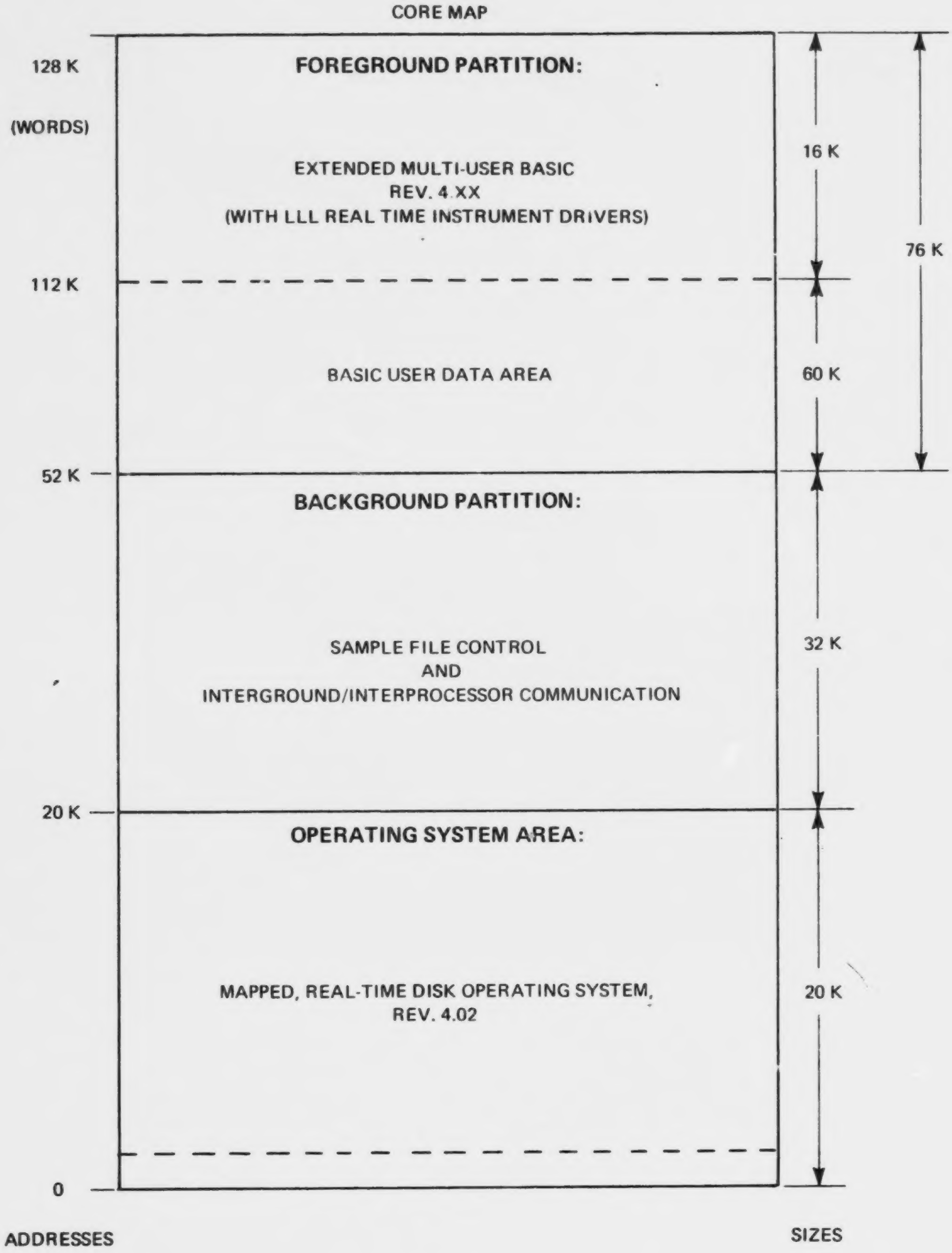

CORE MAP
128 K
(WORDS)
FOREGROUND PARTITION:
EXTENDED MULTI-USER BASIC
REV. 4.XX
(WITH LLL REAL TIME INSTRUMENT DRIVERS)
16 K
76 K
112 K
BASIC USER DATA AREA
60 K
52 K
BACKGROUND PARTITION:
SAMPLE FILE CONTROL
AND
INTERGROUND/INTERPROCESSOR COMMUNICATION
32 K
20 K
OPERATING SYSTEM AREA:
MAPPED, REAL-TIME DISK OPERATING SYSTEM,
REV. 4.02
20 K
0
ADDRESSES
SIZES

Figure 2
Maximum Single Processor Laboratory Automation System Configuration

DATA ACQUISITION SYSTEM FOR AN ATOMIC ABSORPTION SPECTROPHOTOMETER, AN ELECTRONIC BALANCE, AND AN OPTICAL EMISSION SPECTROMETER

By Van A. Wheeler

Computer automation of several instruments within the Analytical Chemistry Branch, Environmental Monitoring and Support Laboratory at Research Triangle Park (RTP), was due primarily to our responsibilities with the National Air Surveillance Network (NASN). The Trace Element Chemistry Section alone was responsible for recording approximately 150,000 data values on file cards each year. One can imagine the possibilities for entry of human error when support of this project required punching the Wang calculator 2 million times a year.

Sample flow of the NASN filter samples is depicted in Figure 1. Eight-by-ten-inch glass fiber filters are screened, numbered, weighed, and then mailed to the field. A 24-hour sample is collected and the operator records collection time, air flow readings, and weather conditions. Upon return of the filter, the final weight is obtained for calculation of total suspended particulate (TSP). Individual filters are cut and combined in a calendar quarterly composite for each site. The acid extract of the composite is analyzed for 24 elements by an optical emission spectrometer with support or additional elemental analysis by atomic absorption spectrophotometry.

The computer system built for the project stores the filter number and initial weight, assigns filters to specific sites, stores the final weight upon receipt from the field, and calculates the TSP based on the site information entered at a CRT terminal at the balance. The system determines the filters to comprise the composite and assigns a sample number to the extract. The analyses are obtained under computer control and the raw instrument data processed with the stored calibration and blank parameters. The extract concentration is merged with previously determined data for each site to produce the final aerometric reports.

The contract to provide such a computer system was finalized with Bendix Field Engineering Corporation in March of 1972. The Automated Laboratory Data Acquisition System (ALDAS), Figure 2, was designed to produce and store valid aerometric data and offer real-time control of three instruments: a Perkin Elmer 403 atomic absorption spectrophotometer (AA) with an automatic sample changer, an Ainsworth 1000D digital balance, and an Applied Research Laboratories 9500 direct reading emission spectrometer (OES). At the time two processors were required as Digital Equipment Corporation's (DEC) real-time operating software could not support both foreground instrument control as well as background report generation and file editing.

Figure 3 demonstrates the hardware organization of the system. The PDP 11/20 foreground processor operates under real-time software and is connected through a bus cable to the instrument interfaces and the CRT's for each instrument. The PDP 11/15 background processor operates under a traditional disk operating system from the 64k word fixed head disk and is connected through a bus cable to a DECTAPE peripheral and line printer. The units on each processor's bus lines are not accessible by the other processor. However, they do share three 1.2 million word disks and the 9-track magnetic tape through a bus switch.

Numerous status checks are performed at each instrument during operation: timing of the sample analysis integration periods are checked with expected values, correct instrument operating modes are checked, and time-out routines are utilized to prevent hangup because of lost data or instrument glitches. After passing preliminary testing, the raw data are recorded on 9-track tape. The tape serves a "log book" function of the analysis recording the date, sample number, and raw data. The raw data are processed according to the instrument's mathematical routines, and the final data are stored on disk file for later report.

The advantages of this system include:

- Turnkey operation of the project. We began with only the analytical equipment and procured a system of all the necessary hardware and software to service the bulk of our NASN analytical responsibilities.
- Elimination of human error in clerical filing and computation.
- Consistency in sample and data treatment.
- Operation of instruments by personnel with little or no experience.

The disadvantages of the system are:

- Failure of the DEC bus switch hardware through which both processors accessed common peripherals at the most inopportune times, requiring a restart of the system and corrupting any open files. Analysis had to be restarted as no allowances for restarting in midstream were provided. The switch was also suspected as the source of ghost messages observed on the AA CRT during maximum use of the three instruments.
- The rigid structure of the system for the NASN project limits its flexibility to process other samples. This "monkey" mode operation is frustrating to professional chemists as there are no allowances for operator decision-making during the analysis scheme other than beginning and ending the program.

The inflexibility of the system has proven to be the death of the system before it could prove its merit. One of the reorganizations within EPA transferred most NASN responsibilities to regional offices leaving the Analytical Chemistry Branch with receipt of a portion of the filter for analysis. The ALDAS software could not support this change in procedure without major revisions.

The system is currently being modified (Figure 4) with the emphasis on servicing the analytical instrumentation instead of a specific project. DEC now has available a real-time operating system, RSX-11M, which can accomplish with one CPU the foreground-background duties desired. Thus the PDP 11/15 and bus switch have been eliminated from the system.

Memory capacity of the PDP 11/20 can be increased from 32k to 128k through a field modification and upgrade to a PDP 11/40. Increased memory will be necessary for real-time operation of anticipated instrument additions to the system since the instruments are so designed that they must be monitored continuously. Thus, it is cheaper to purchase additional memory than to modify existing instrumentation to take advantage of the time-sharing capabilities of the computer operating system.

Software for the new system will be written in two operating modes. A routine or "monkey" mode much like the original ALDAS concept will be maintained where standards and quality control samples must meet rigid historical limits. The second mode will be for special samples with the operator having considerable input as to selection of samples, standards, and quality control samples. In this mode, the software will evaluate the standards and quality control data generated during the analysis and report the statistical variation of the analysis. The log tape concept will be maintained, but the data will be written on a DECTAPE unit for faster access.

In conclusion, automation of our analytical laboratory is now being approached from the standpoint of servicing the analytical equipment instead of a specific project. From this position, the outermost layer of programing can reflect project needs. If program plans change, minimal, if any, programing changes will be necessary.

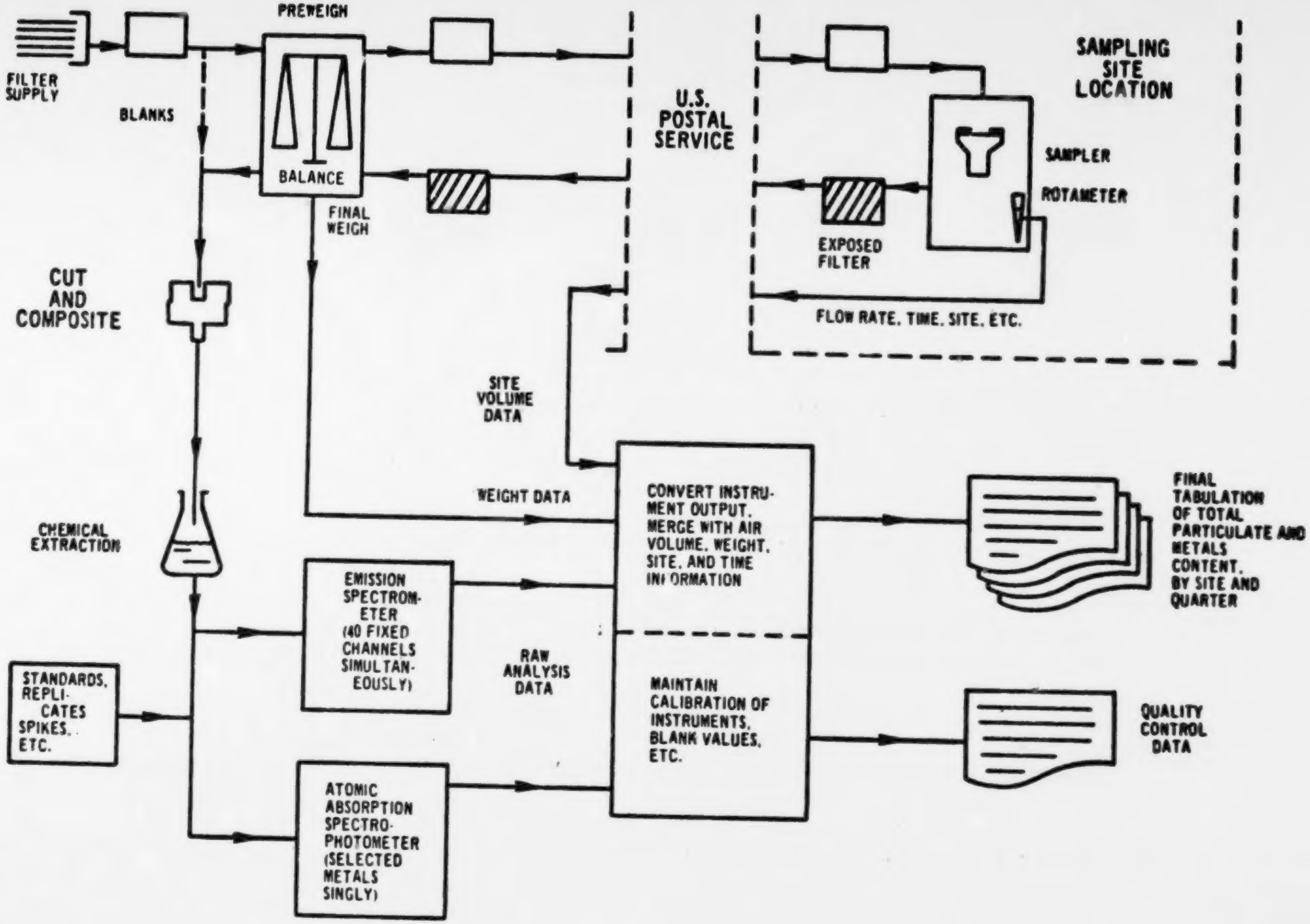

FILTER SUPPLY
BLANKS
PREWEIGH
BALANCE
FINAL WEIGH
U.S. POSTAL SERVICE
SAMPLING SITE LOCATION
SAMPLER
ROTAMETER
EXPOSED FILTER
FLOW RATE, TIME, SITE, ETC.
CUT AND COMPOSITE
SITE VOLUME DATA
WEIGHT DATA
CHEMICAL EXTRACTION
EMISSION SPECTROMETER (40 FIXED CHANNELS SIMULTANEOUSLY)
STANDARDS, REPLICATES SPIKES, ETC.
RAW ANALYSIS DATA
ATOMIC ABSORPTION SPECTROPHOTOMETER (SELECTED METALS SINGLY)
CONVERT INSTRUMENT OUTPUT, MERGE WITH AIR VOLUME, WEIGHT, SITE, AND TIME INFORMATION
MAINTAIN CALIBRATION OF INSTRUMENTS, BLANK VALUES, ETC.
FINAL TABULATION OF TOTAL PARTICULATE AND METALS CONTENT, BY SITE AND QUARTER
QUALITY CONTROL DATA

Figure 1
NASN Filter and Data Processing

PURPOSE:	TO PRODUCE AND STORE VALID AEROMETRIC LEVELS OF TRACE ELEMENTS AND TOTAL SUSPENDED PARTICULATE MATTER IN REAL TIME
INSTRUMENTS:	1. PERKIN-ELMER 403 AUTOMATIC ABSORPTION SPECTROPHOTOMETER 2. AINSWORTH 1000D DIGITAL BALANCE 3. ARL 9500 DIRECT READING SPECTROMETER
PROCESSORS:	1. DEC PDP 11/15 - BACKGROUND (24K MEMORY) REPORT GENERATION 2. DEC PDP 11/20 - FOREGROUND (12K MEMORY) INSTRUMENT DATA PROCESSING
INPUT:	1. CRT TERMINALS (FOUR) - DESCRIPTIVE AND NONINSTRUMENTAL DATA 2. ANALYTICAL INSTRUMENTS (THREE) - INSTRUMENT RESPONSES
OUTPUT:	1. 9-TRACK MAGNETIC TAPE - FINAL PRODUCT 2. LINE PRINTER - PRINTED RECORDS, REPORTS 3. MAGNETIC DISK PACK (1.2 MILLION WORDS) - SOURCE OF PERMANENT FILES 4. CRT TERMINALS - TEMPORARY DISPLAY

Figure 2
ALDAS Overview

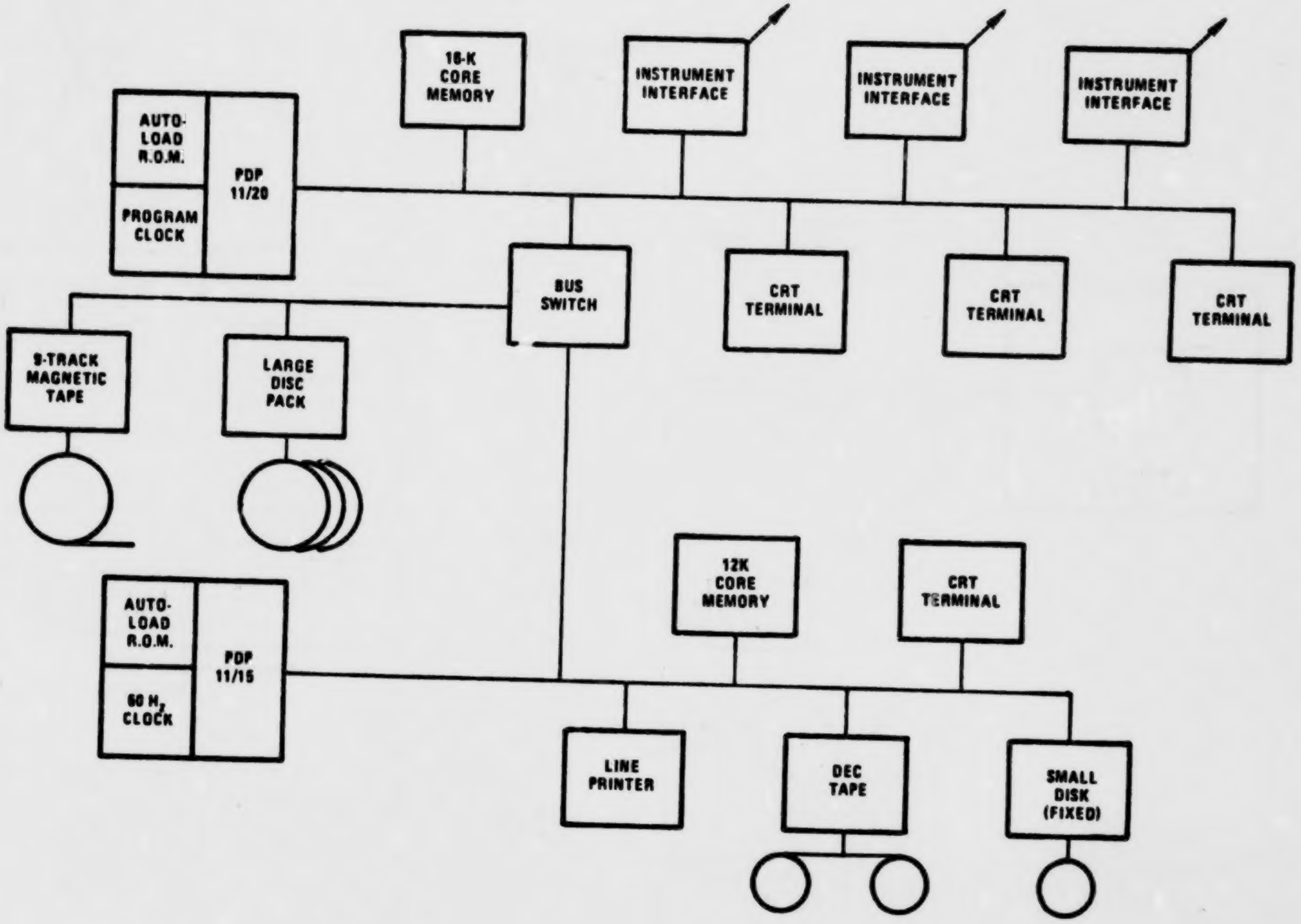

AUTO-LOAD R.O.M.
PROGRAM CLOCK
PDP 11/20
16-K CORE MEMORY
INSTRUMENT INTERFACE
INSTRUMENT INTERFACE
INSTRUMENT INTERFACE
BUS SWITCH
CRT TERMINAL
CRT TERMINAL
CRT TERMINAL
9-TRACK MAGNETIC TAPE
LARGE DISC PACK
AUTO-LOAD R.O.M.
60 H_z CLOCK
PDP 11/15
12K CORE MEMORY
CRT TERMINAL
LINE PRINTER
DEC TAPE
SMALL DISK (FIXED)

Figure 3
Computer Hardware

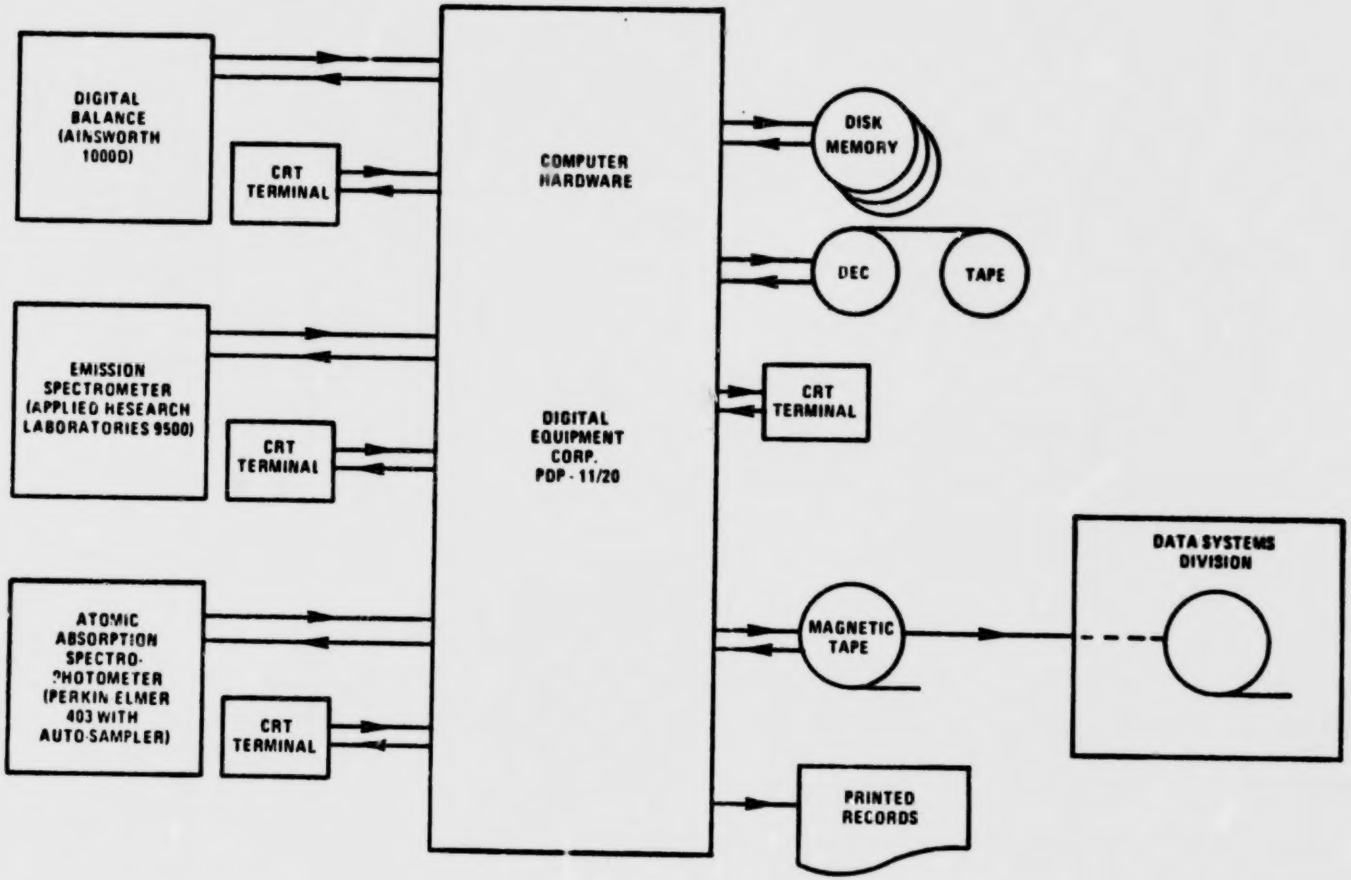

DIGITAL BALANCE (AINSWORTH 1000D)
CRT TERMINAL
EMISSION SPECTROMETER (APPLIED RESEARCH LABORATORIES 9500)
CRT TERMINAL
ATOMIC ABSORPTION SPECTRO-PHOTOMETER (PERKIN ELMER 403 WITH AUTO-SAMPLER)
CRT TERMINAL
COMPUTER HARDWARE
DIGITAL EQUIPMENT CORP. PDP - 11/20
DISK MEMORY
DEC
TAPE
CRT TERMINAL
MAGNETIC TAPE
DATA SYSTEMS DIVISION
PRINTED RECORDS

Figure 4
ALDAS System

AN AUTOMATED ANALYSIS AND DATA ACQUISITION SYSTEM

By Michael D. Mullin

The Large Lakes Research Station is the EPA facility charged with research into the fate and transport of pollutants in large lakes, concentrating on the Great Lakes. A project was initiated 2½ years ago to study Saginaw Bay in conjunction with EPA and other Agency-sponsored studies on Lake Huron.

Initially the Saginaw Bay study entailed 59 stations in the Bay to be sampled at various depths for a total of approximately 110 samples per cruise. As a result of knowledge gained during the 1974 field season, the number of stations was decreased to 37 for a total of 80 samples per cruise for the 1975 season. Each sample was analyzed in the laboratory for 15 to 25 parameters, including nutrients, organic carbon, conservative and trace metals, chlorophyll, and chloride. At the same time, other research groups were conducting extensive biological studies to interact with our physical and chemical data. There was a total of 31 cruises amounting to over 50,000 separate analyses.

Within budgetary constraints, it was necessary to automate as many of the analyses as practicable. For this purpose, a Technicon Auto-Analyzer II System was purchased to, initially, provide the analyses of five nutrients: dissolved ammonia, dissolved reactive silicate, dissolved reactive phosphates, dissolved nitrate and nitrite, and dissolved sulfate. Four additional parameters have since been added: total phosphorus, total kjeldahl nitrogen, chloride, and dissolved hexavalent chromium. Dissolved iron may be added to the system in the near future.

Although there are many other parameters being analyzed in the laboratory, the only additional ones to be automated analogous to the Technicon discrete sampling technique are sodium, potassium, calcium, and magnesium by atomic absorption spectrometry. With our present dual channel instrument, calcium and magnesium are analyzed simultaneously. Due to different burner conditions, sodium and potassium must be analyzed separately.

With 10 parameters, this system as shown in Figure 1 can generate a large volume of data. For an effective workday of 6 hours, up to 1,700 results can be generated for the 10-parameter system per day. An additional 500 results per day can be generated by the dual channel automated atomic absorption spectrometer. The Auto-Analyzer System can be set up and maintained by two or three technicians. In addition, if they have to read peak heights from a strip chart recorder, perform regression analyses on standards, and calculate concentrations of samples, more time will be spent on calculations than on analyzing samples. The other option is to have an online data acquisition system mated to the instruments for time-consuming work.

Prior to purchasing any automated data processing equipment, a requirement for our laboratory was the need for the system to be functioning with as little in-house effort as possible. The decision was made to purchase a Digital Equipment Corporation PDP-8e minicomputer. The basic unit is a 12-bit digital computer with 8k core memory, 12-channel analog multiplexer, and a teletype for communication and printing of results. This has since been expanded to 32k of core, along with a high-speed reader-punch, a medium-speed line printer, and a punched card reader.

There were a few problems that required attention prior to online operation. The first priority item was getting the analog signal from the analytical instrument to the computer. The voltage output from the Auto-Analyzer colorimeter is 0 to +5 volts DC, and the required analog input to the analog to digital convertor is -1 to +1 volts DC. To eliminate this problem, an interface was designed and constructed utilizing an operational amplifier and several resistors to produce this conversion and is shown in Figure 2. The system required that the output be linear over the entire operating range and that the final voltage be adjustable so as to fine tune the signal.

The analog output of the colorimeter is the same as that going to the strip chart recorder. At the low concentration levels encountered with some of the parameters, there is a slight drift in the baseline absorbance that must be taken into account when calculating standard graphs or sample concentrations. Digital Equipment Corporation developed a computer language analagous to FORTRAN called FOCAL. For the PDP-8, they market a software package called PAMILA, which is an overlay to FOCAL and modifies it. We subsequently adapted PAMILA to our specific uses and needs for real-time operational control.

Figure 3 illustrates the baseline drift correction procedure which is part of the in-house modified PAMILA package. The routine checks the between-peak valleys. If one falls below the initial baseline, the baseline is reset and the peak heights are then corrected by subtracting the interpolated baseline. If the baseline increases by the end of a preset time interval, the baseline is reset and the peak heights proportionally corrected over the time interval. A number of other procedures also yield relative peak area, time of peak ending, peak height, and type of peak.

The 8k version can hold information only for 64 peaks in the peak file storage at any one time, so a mechanism is provided for printing of the contents of the buffer, by channel, at any preset time interval. The time interval is set for every ten sample cups. The first and last sample in each decade is a water wash to set the initiating and terminating baselines for the computer. With 12 channels online simultaneously the buffer possibly could have up to 192 peaks stored at one time. With an additional 8k or more of core, the peak file storage can hold information on 200 peaks, thus increasing the number of channels that can be monitored at any one time.

Figure 4 shows the raw data output format. A punch paper tape copy of the raw data is generated simultaneously with the typed copy. When a given series of samples, usually a day's run, is completed, the punched tape is fed into the computer and stored on the disk prior to editing and analyzing. Then the raw data can be recalled, and extraneous peaks, noise, or other unwanted information deleted. The corrected raw data file is then analyzed with an interactive program developed at the Grosse Ile Laboratory to give concentrations to the standard peak heights, set upper and lower limits, perform a regression on the standards and, using this equation, calculate the concentrations for the unknowns. The output has all the needed regression factors together with the calculated concentrations. Figure 5 is a copy of the output.

There are a number of additional ways to utilize the computer. Data management procedures can be implemented. Additional instruments, such as pH meters, carbon analyzers, and weighing balances, can be interfaced with the existing equipment. Essentially, any laboratory instrument that generates a measurable signal will provide this interface. The Large Lakes Research Station hopes to implement some of these as time and resources permit.

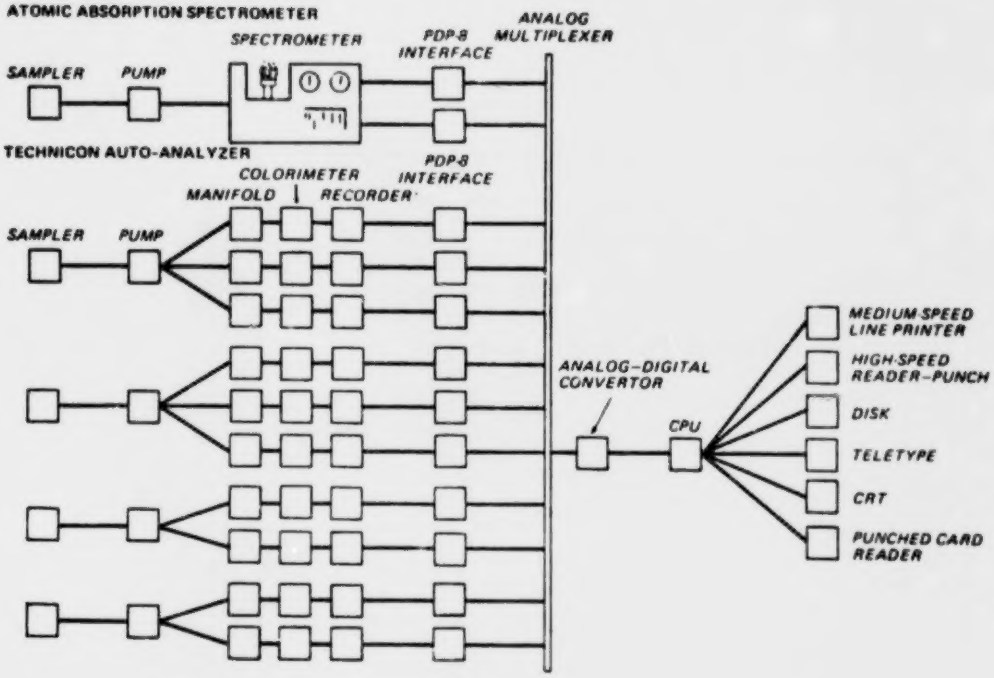
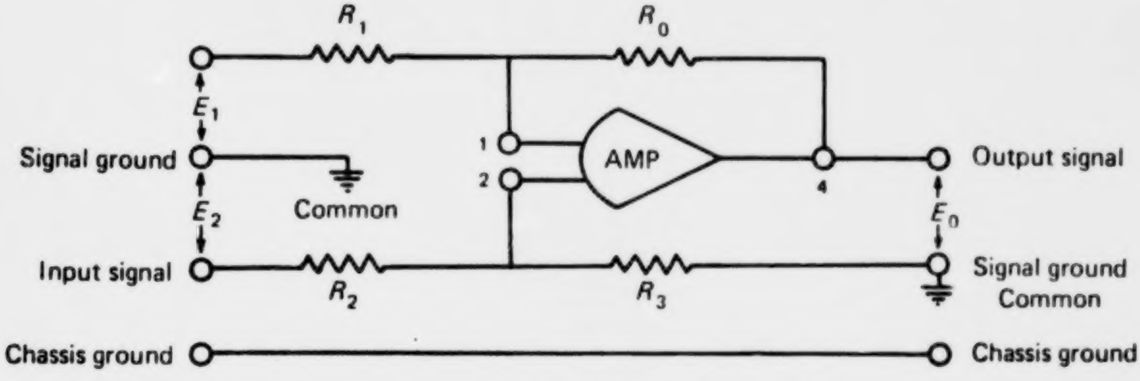

ATOMIC ABSORPTION SPECTROMETER
SPECTROMETER
PDP-8 INTERFACE
ANALOG MULTIPLEXER
SAMPLER
PUMP
TECHNICON AUTO-ANALYZER
PDP-8 INTERFACE
COLORIMETER
MANIFOLD
RECORDER
SAMPLER
PUMP
ANALOG–DIGITAL CONVERTOR
CPU
MEDIUM-SPEED LINE PRINTER
HIGH-SPEED READER–PUNCH
DISK
TELETYPE
CRT
PUNCHED CARD READER

Figure 1
Twelve Channel Automated System


R_1
R_0
E_1
Signal ground
E_2
Common
Input signal
R_2
1
2
AMP
4
R_3
Output signal
E_0
Signal ground
Common
Chassis ground
Chassis ground

R_0 = 10 K

R_1 = 50 K*

R_2 = 20 K*

R_3 = 10 K

AMP = 3267/12C Burr–Brown

$$E_0 = \left(\frac{R_3}{R_1}\right)\left(\frac{R_1 + R_0}{R_2 + R_3}\right)E_2 - \left(\frac{R_0}{R_1}\right)E_1$$

E_1 = 5 VDC

---

*Variable potentiometer.

Figure 2
Analytical Instrument PDP-8 Interface

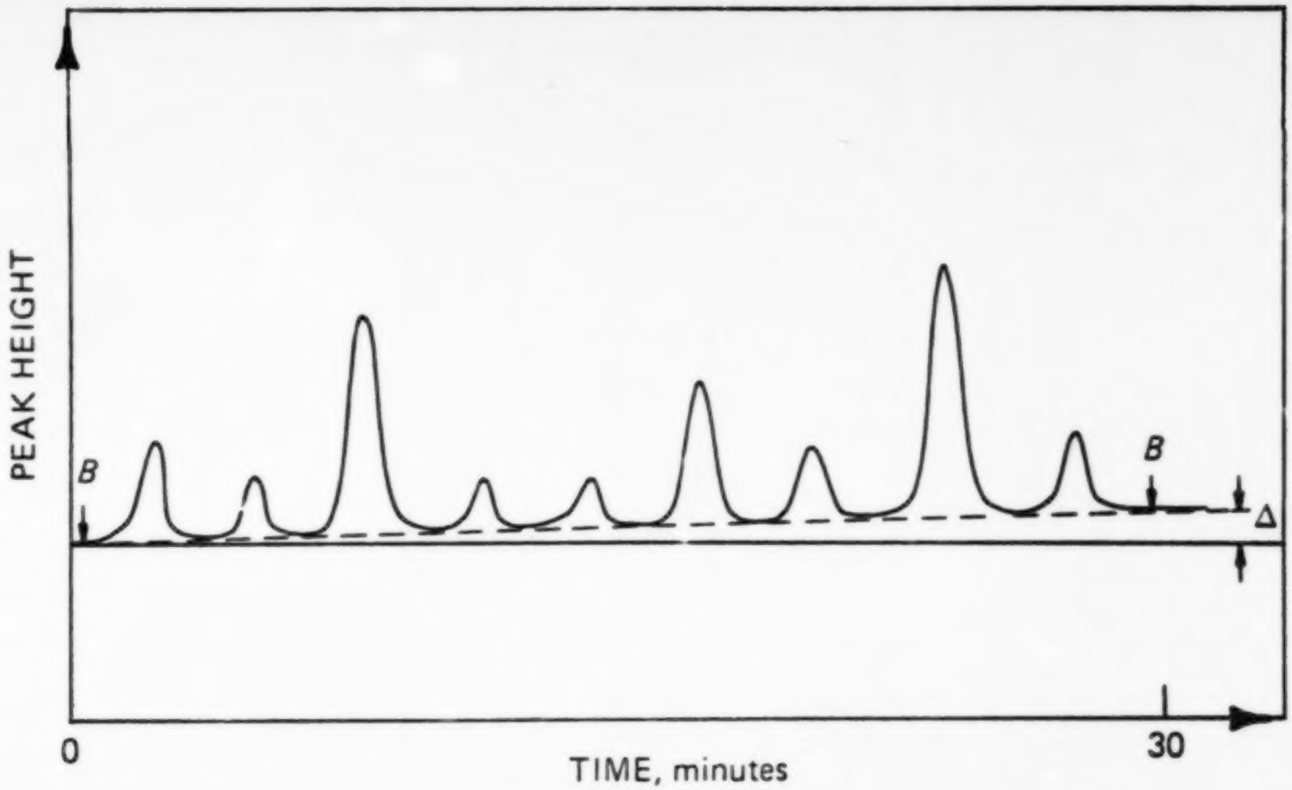


Figure 3
Baseline Connection

```
                                                                    AT:     1
RUN #     2  INST:      1  DAY    11  TIME:  10:    6

     #          AREA             RET.T         HGT   TYPE

     1        0.91783            3.683          20    BV
     2        1.77440            7.167          42    VV
     3        5.20730            10.23         123    VV
     4        5.27319            13.23         126    VV
     5        8.65888            16.37         204    VB
     6        17.5442            19.50         422    BB
     7        25.6064            22.57         618    BB
     8        35.0179            25.50         834    BB

TA   0.215524E+06
```

```
                                                                    AT:     5
RUN #     3  INST:      0  DAY    11  TIME:  10:    6

     #          AREA             RET.T         HGT   TYPE

     1        0.98586            3.883          19    BV
     2        2.03330            7.350          42    VV
     3        5.56292            10.62          95    VV
     4        6.20325            13.42          96    VV
     5        9.32638            16.55         161    VV
     6        16.7614            19.55         353    VB
     7        25.0732            22.55         543    BB
     8        34.0537            25.48         731    BB

TA   0.185625E+06
```

```
                                                                    AT:     3
RUN #     4  INST:      2  DAY    11  TIME:  10:    6

     #          AREA             RET.T         HGT   TYPE

     1        1.18331            3.700          24    BV
     2        1.76642            7.333          41    VB
     3        5.11170            10.60         119    BB
     4        5.28196            13.53         123    BB
     5        8.55670            16.53         198    BB
     6        18.0754            19.60         418    BB
     7        25.7231            22.60         588    BB
     8        34.3015            25.60         789    BB

TA   0.211437E+06
```

Figure 4
On-Line Raw Data Output

```
                              ***** ANALYSIS TYPE =  3 *****

      ACTUAL         ACTUAL       CALCULATED     CONCENTRATION

   PEAK HEIGHTS     STANDARDS      STANDARDS     DIFFERENCES
        24.          0.00125        0.00125       -0.00000
        41.          0.00250        0.00233        0.00017
       119.          0.00750        0.00728        0.00022
       123.          0.00750        0.00753       -0.00003
       198.          0.01250        0.01229        0.00021
       418.          0.02500        0.02625       -0.00125
       588.          0.03750        0.03703        0.00047
       789.          0.05000        0.04978        0.00022

MINIMUM DETECT CONCENTRATION =     0.00030

MAXIMUM DETECT CONCENTRATION =     0.05000

DEGREE OF POLYNOMIAL =  1

CALCULATED CONCENTRATION =      -0.0002673    +       0.0000634 *PEAK HEIGHT FOR LEAST SQUARES

CORRELATION COEFFICIENT =  0.9995630

CALCULATED CONCENTRATION =      -0.0002673 +       0.0000634 *PEAK HEIGHT. FOR POLFIT

REDUCED CHI SQUARE FOR FIT =  0.0000003


RUN #  7 INSTR #  2 DAY 11 TIME 10:36 AT:  3
          RETT       HGT          XCONC
   1      3.067      77.          0.0046
   2      7.500      98.          0.0059
   3     10.500       5.      LESS THAN:  K      0.0003
   4     13.230      75.          0.0045
   5     16.230       1.      LESS THAN:  K      0.0003
   6     19.300      61.          0.0036
   7     22.570      62.          0.0037
   8     25.170      32.          0.0018
```

Figure 5
Off-Line Calculated Concentration Output

A TURNKEY SYSTEM: THE RELATIVE ADVANTAGES AND DISADVANTAGES

By D. Craig Shew

The primary objective of this paper is to discuss some of the relative advantages and disadvantages of automated laboratory instrumentation in terms of commercially available turnkey systems based on our past 4 years' experience with an automated gas chromatography/mass spectrometry (GC/MS) system. It is important to point out that, in general, the people who either operate or are responsible for EPA's mass spectrometry laboratories are trained in organic or in analytical chemistry rather than in data processing, per se. Thus, this discussion will be from the viewpoint of the end user of specific instrumentation as opposed to that of one who is involved in general hardware and software design.

About 4 years ago, EPA made a major commitment to the automated GC/MS system by purchasing 23 more or less similar systems at a cost of about $2.5 million. In general, the GC/MS system provides an unequivocal basis for the identification of organic compounds and, thus, finds many uses in organic analytical chemistry. Specifically, EPA uses mass spectrometry for the identification of organic pollutants originating from a wide variety of sources. At this particular point in the evolution of automated instrumentation, the GC/MS system has proved to be one of the most widely used, most successful examples to date.

Figure 1 shows an overview of the system with its three main components: a gas chromatograph, mass spectrometer, and data system with the various input-output devices. The system is controlled by a DEC PDP-8 minicomputer which provides for control of the mass spectrometer, data acquisition, data manipulation and reduction, and data output. The entire system was purchased from the Finnigan Corporation at a cost of about $90 thousand. The data system was developed, in part, and constructed by Systems Industries, a small captive subcontractor to Finnigan Corporation, at a cost of about $55 thousand. Since that time, the Finnigan Corporation has developed its own data system and is now in direct competition with Systems Industries. As a result, we are in the rather tenuous position of having to depend on the Finnigan Corporation for service of the mass spectrometer and having to depend on Systems Industries for service and support of the data system. Although this is an undesirable position from the standpoint of service responsibility, no major problems have arisen so far.

One of the first advantages of a commercially available turnkey system is the lower cost when compared to the various alternatives. Development and update costs are amortized over a large number of users, thus making the turnkey system more economical. A major advantage from the standpoint of the end user is that the system is available for immediate use at a fixed cost. Thus, the system's capabilities and limitations can be compared with other techniques or with other commercially available systems.

Another major advantage for EPA's MS laboratories is a well-organized, active users' group, a direct result of having 23 similar systems within the Agency. Consequently, a number of people have contributed to the development of standardized operational and analytical techniques. This has saved a substantial duplication of effort in that development of similar techniques was not required for various configurations of similar instrumentation. In terms of day-to-day operations, the 23 systems together have substantially lessened the degree of expertise needed in any one laboratory because of support from the other users. Additional advantages have resulted from having similar commercial systems and an active users' group within the Agency. For example, data and software changes can be easily analyzed or evaluated by other laboratories with similar hardware. Similarly, quality control is easier to set up and maintain.

One of the major disadvantages of the commercially available turnkey system is that one becomes highly dependent on others for hardware service and software updating and support. However, it might be well to point out that primarily because of the commercially competitive situation, about ten software updates and revisions have been obtained at essentially no cost to us. Hiring of outside contractors to make specific software changes and additions in some cases has been a fairly expensive proposition. A second disadvantage is that software listings are generally unavailable to the commercial users; consequently, even minor software changes are difficult to accomplish. In our particular case, we have able to obtain software listings, but the documentation has been so poor that the listings are practically useless.

In considering the various alternatives to commercially available turnkey systems, there are several options available. Some form of interagency agreement is probably the most attractive since generally the contract is relatively simple, and there is no question of ownership rights. Similarly, a system can be developed on a commercial basis according to detailed specifications outlined in a contract. Also, one can envision some sort of in-house effort in which commercially available hardware and software are assembled, or alternatively, a more extensive effort, including the necessary research and development, construction of hardware, and writing of software. The latter case involves a multidisciplinary effort and is generally beyond the capabilities of most EPA laboratories.

In conclusion, our experience with a commercially available GC/MS system has been very satisfactory. However, in other cases of specific types of automated instrumentation, the various alternatives to a turnkey system would have to be considered on a point by point basis.

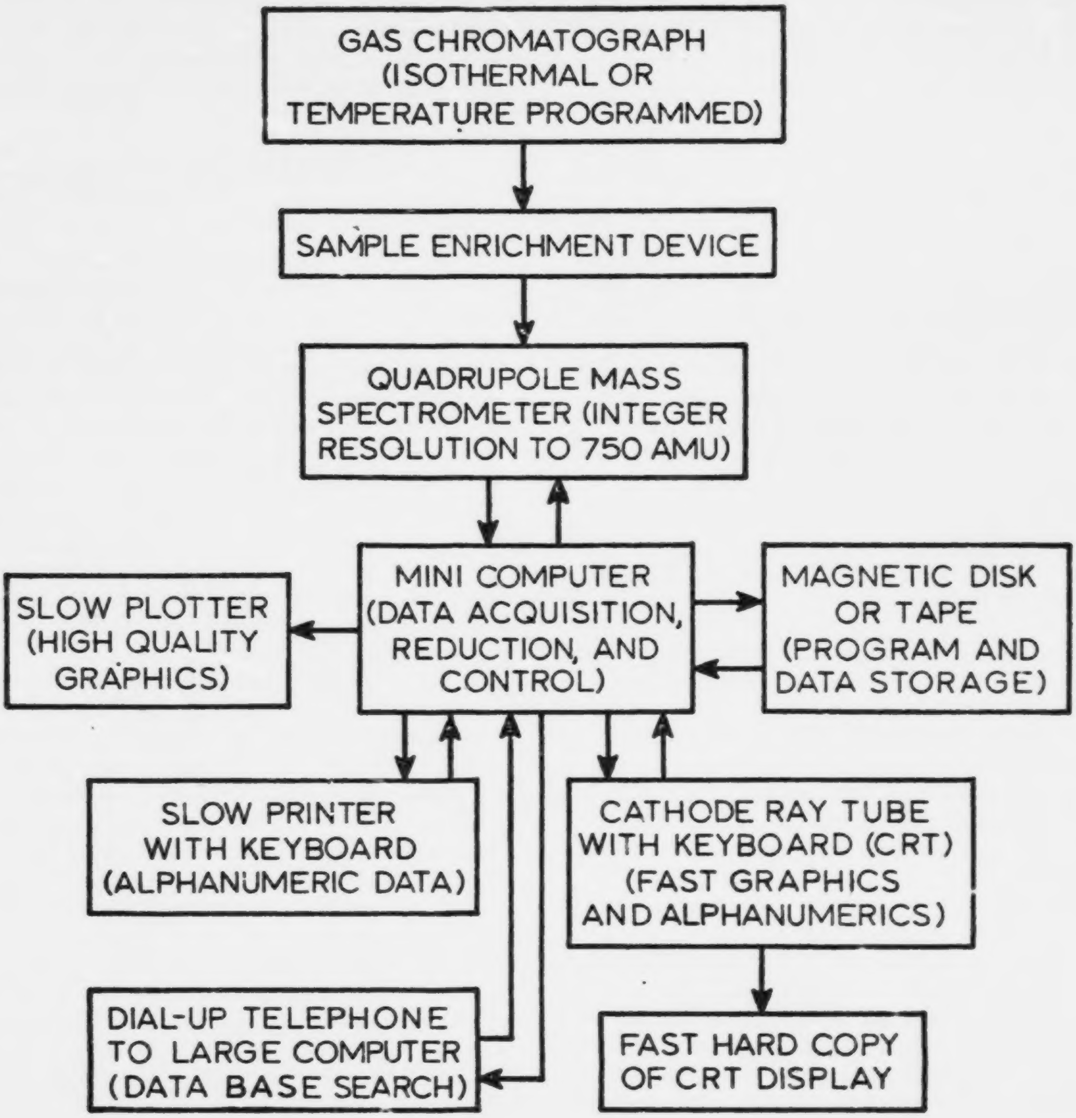

GAS CHROMATOGRAPH (ISOTHERMAL OR TEMPERATURE PROGRAMMED)
SAMPLE ENRICHMENT DEVICE
QUADRUPOLE MASS SPECTROMETER (INTEGER RESOLUTION TO 750 AMU)
SLOW PLOTTER (HIGH QUALITY GRAPHICS)
MINI COMPUTER (DATA ACQUISITION, REDUCTION, AND CONTROL)
MAGNETIC DISK OR TAPE (PROGRAM AND DATA STORAGE)
SLOW PRINTER WITH KEYBOARD (ALPHANUMERIC DATA)
CATHODE RAY TUBE WITH KEYBOARD (CRT) (FAST GRAPHICS AND ALPHANUMERICS)
DIAL-UP TELEPHONE TO LARGE COMPUTER (DATA BASE SEARCH)
FAST HARD COPY OF CRT DISPLAY

Figure 1
GC/MS System

ONE APPROACH TO LABORATORY AUTOMATION

By Jack W. Frazer

It has been said often that an antiphilosophical attitude exists in American society. While such attitudes are considered normal, ours, it is said, is intensified as a result of our bias towards action.[1] The effects of our bias towards action combined with our antiphilosophical attitude are nowhere more apparent than in our efforts to automate laboratories and chemical processes. Most of this kind of automation has proceeded via the action route with too little thought given to an understanding of the many dimensions involved and how the associated efforts might be accomplished most effectively.

A typical scenario is as follows: development of vague ideas concerning instruments and the functions to be automated, survey of the computer market, selection and purchase of a computer and auxiliary hardware, acceptance of the computer and accessories, and finally, an *attempt* to build a suitable system without the aid of specifications or well-defined plans. This approach has resulted in many systems that have failed to produce significant cost savings or improved scientific results. Note that the usual procedure is based on action; that is, purchase of a computer first without the aid of a complete set of system specifications and plans (an unacceptable philosophy).

There are undoubtedly many philosophies that, when put in practice, could result in the successful and cost-effective implementation of online computer automation. Following is an outline of one philosophy as set forth at the Lawrence Livermore Laboratory,[2-6] which is now being fully developed and documented by Committee E-31 of ASTM.[7-10]

One of the first tenets of this philosophy is: Computers used for online automation of instruments and processes are system problems. Often the computer presents the designers with the least difficult problems. If, then, the computer is not always the central issue in the development of computer automation, what are the important issues and dimensions? First, the designer must recognize that an automated laboratory impacts many aspects of management practices, chemical and physical processes being automated, and the instrument operational procedures and characteristics. Therefore, information from a number of people with different responsibilities and expertise is required in order to properly define and specify the desired characteristics of the proposed automation.

Secondly, the designer should recognize that the implementation of an automated laboratory is a problem of many dimensions. Therefore, an interdisciplinary team effort is required if the implementation is to proceed smoothly and in a cost-effective manner. Due to the complexity of automation and the requirement for multidisciplinary team action, larger automation projects are difficult to manage. This is particularly true when the implementation is undertaken with incomplete specifications and designs.

OPERATIONAL PROCEDURES

Since automation is recognized as a difficult and complex undertaking requiring effort from many scientific and engineering disciplines, why not attack these projects as we do other difficult tasks; that is, separate the variables and solve them one at a time? An operational procedure that has been field tested and found to be effective is shown in Table I.

Table I
Operational Procedures

- System definition (including a cost benefit analysis)
- System specifications
- Functional design
- Implementation design (hardware and software selection)
- System implementation
- System evaluation
- Documentation

System Definition

At the onset of an automation project, the responsible scientist should write a brief tutorial description of the proposed project aimed at those levels of management responsible for funding. Therefore, it should contain only a brief description of the principal features of the project and the anticipated benefits. For larger systems, a schematic representation should be included. Finally, when the system specifications and functional design are complete, a cost benefit analysis should be inserted into the system definition.

System Specifications

System specifications may be defined as a listing of all the details necessary to direct the uninformed (but knowledgeable scientist) in the construction, installation, and testing of a complex project. They are similar in detail and extent to the specifications required to build a complex instrument, dam, or factory. For automation projects, we have preferred to consider the specifications to exist in one of three domains; inputs, outputs, and transfer functions. The following is a set of definitions for these terms:

- Inputs - any source of stimuli that causes a response within the system
- Outputs - actions taken by the system as a result of stimuli
- Transfer functions - those algorithms that interconnect the system inputs and outputs.

One of the better ways to understand system specifications is to study briefly a case history as published in ASTM STP-578. Included below are examples taken from one of these papers.[10] The system has now been delivered. Tables 2 and 3 contain a few of the input specifications for one of four atomic absorption instruments in the system.

Table 4 is an example of output specifications of an output report.

Figure 1 shows part of one transfer function specification for the atomic absorption instruments. It describes the control algorithm for the automatic samplers.

The above examples are only a small part of the system specifications, which become formidable documents running to several hundred pages for large systems.

Functional Design

A functional design is a schematic representation of an automated system, including inputs, outputs, and the interconnecting transfer functions. An example of a functional design is shown in Figure 2 for Auto Analyzers automated with the other instruments described previously. Where there are severe system time-response or bandwidth requirements, the required time-response characteristics and data rates are listed on all data paths.

Implementation Design

After the above three phases of work are completed, the designers begin the selection of specific hardware and software as required to meet the design requirements. The cost effectiveness of various hardware-software tradeoffs, as well as computer operating system (software) characteristics, are assessed. With the aid of the system specifications, it is a relatively straightforward procedure to select system components that will meet performance requirements.

CONCLUSION

Given a good set of system specifications, an implementation design, and appropriate hardware and software, it is usually a fairly straightforward task to construct the computer automated system. However, no system is really complete until it is fully documented and evaluated. Evaluation should include not only the proper set of tests to assure that the system meets design specifications but also the tests that determine the "boundary conditions" of the system. These include such parameters as the system time-response and bandwidth under various operating conditions.

REFERENCES

1 Tinder, Glenn, "Political Thinking," Little, Brown and Co., Boston 1974.

2 Frazer, Jack W., "Management of Computer Automation in the Scientific Laboratory," UCRL-72162. Presented at Stored Program Controller Symposium, Sandia Laboratory, Albuquerque, New Mexico, September 23, 1969.

3 Frazer, Jack W., "Laboratory Computerization: Problems and Possible Solutions." Presented at ASTM Meeting, Philadelphia, Pennsylvania, May 12, 1970.

4 Frazer, Jack W., "A Systems Approach for the Laboratory Computer Is Necessary," *Materials Research Standards*, vol. 12, no. 2 (1972), pp. 8-12.

5 Frazer, Jack W., "Design Procedures for Chemical Automation," *Amer. Lab.*, vol. 5, no. 2 (1973), pp. 39-49.

6 Frazer, J. W., Perone, S. P., and Ernst, K., "A Systematic Approach to Instrument Automation," *Amer. Lab.*, vol. 5, no. 2 (1973), pp. 39-49.

7 Frazer, J. W. and Kunz, F. W. editors, "Computerized Laboratory Systems," *ASTM Special Technical Publication 578*, ASTM, 1975.

8 Frazer, J. W., Kray, A. M., Boyle, W. G., Morris, W. F. and Fisher, E., "The Need for Automated System Specifications and Designs," *ASTM Special Technical Publication 578*, ASTM, 1975, pp. 65-76.

9 Frazer, J. W., Perone, S. P., Ernst, K., and Bland, H. R., "Recommended Procedure for System Specification and Design: Automation of a Gas Chromatograph-Mass Spectrometer System," *ASTM Special Technical Publication 578*, ASTM, 1975, pp. 25-64.

10 Frazer, J. W. and Barton, G. W. Jr., "A Feasibility Study and Functional Design for the Computerized Automation of the Central Regional Laboratory, EPA Region V, Chicago," *ASTM Special Technical Publication 578*, ASTM, 1975, pp. 152-256.

Table 2
Some of the "Operator and File Inputs" Specifications for the Atomic Absorption Instruments

Inputs for Operator Interactions During the Course of a Run

1. Command to halt a run in the event of out-of-control conditions or equipment failure

2. Command to restart the run in the event of an interruption. This would include:

   (a) Sampler position.
   (b) Identification of the starting solution

3. Commands to set pause time and integration time when in the flame aspiration mode.

4. Analysis commands for the semiautomatic sampler mode of operation:

   (a) S = standard, S1∧N = sample standard 1 Nth reading.
   (b) U = unknown, U1∧N = sample unknown 1 Nth reading.
   (c) B = baseline.
   (d) A = average the replicates.
   (e) E = erase, S2∧2E, erase 2nd replicate run of standard 2.
   (f) C = calculate, C1 = interpolation, C1, C2 = first, or second degree least squares fit.
   CA = method of addition.

5. Special commands for the preprocessed sample mode of operation (for example, graphite furnace):

   (a) INT! or PK! for integration or peak height value.
   (b) PN∧N!, read Nth peak (for example PH∧3! read the third peak).
   (c) U1∧S1∧N; unknown 1 + standard 1, Nth reading (for standard additions).
   U1∧S2∧N
   U5∧S3∧N; unknown 5 + standard 3, Nth reading.

6. Quality control commands:

   (a) SC5 = check standard as standard 5 run as a check,
   SC5∧N = check standard 5, Nth reading.
   (b) SP5 = spiked unknown 5, SP5∧N = Nth reading of spiked unknown 5.

7. Reagent blank commands:

   (a) Y = use reagent blank values to correct results.
   (b) N = do not use reagent blank values to correct results.
   (c) ? = skip this value for now.
   (d) AV = Use average value of reagent blanks.

Table 3
Some of the "Instrument Inputs" Specifications for the Atomic Absorption Instruments

Instrument Inputs

General

1. Dynamic range of signal: 10,000 for -5 fullscale: 500 μV must be detectable.
2. Example of signal: Appendix I-K has an example of a typical AA signal.

PE 303

The PE 303 will be fitted with a PE DCRI readout which will have the following electrical characteristics:

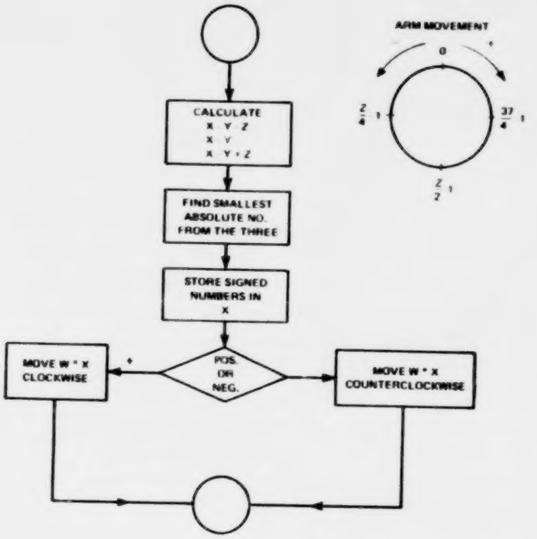
1. Signal characteristics:
    - (a) Internal sources -5 V fullscale from a solid-state operational amplifier output impedance of < 10 Ω.
    - (b) Sample source 0 to -5 V fullscale. Pin 9 of the interface board in the Model DCRI readout.
    - (c) Reference source -3 ± 1.5V. The exact value is dependent upon the energy level of the source. Pin 11 of the Interface Board in the Model DCRI readout.
    - (d) Noise 50 mV of 1 μs pulses dependent upon the noise-suppression switch setting. Time constants 2 to 80 s in 5 settings.
2. Required filtering:

    Analog; low pass filter; 4 pole rolloff.
    Suggested at 0.1, 1, 10Hz.

Table 4
Example of an Output Report for an Electronic Balance in the Automated System

Example of Listings

```
/ = Explanations and comments          ! = Carriage return, line feed.
/  Underlined responses are operator's input.
?  RUN SFC!                            /  Call sample file controller.
   SUBSCHEMA?      FILTER!

?  LIST GPDATE; GPST; EAST CHICAGO; FILTER; NETWT!
   PASSWORD?       LOGSDON!

/  Display the net weight of all the filters.
/  By date from East Chicago.

GPDATE         GPST          FILTER      NETWT

 740921     EAST CHICAGO    4571293      45.230    /gram
                            4571294      43.211
                            4571295      49.832
   .                           .            .
   .                           .            .
   .                           .            .
 741021     EAST CHICAGO    5678912      48.213
                            5678913      48.214
                               .            .
                               .            .
                               .            .
```



ARM MOVEMENT
0
+
Z/4 - 1
3Z/4 - 1
Z/2 - 1
CALCULATE
X - Y - Z
X - Y
X - Y + Z
FIND SMALLEST ABSOLUTE NO. FROM THE THREE
STORE SIGNED NUMBERS IN X
POS. OR NEG.
+
MOVE W * X CLOCKWISE
MOVE W * X COUNTERCLOCKWISE

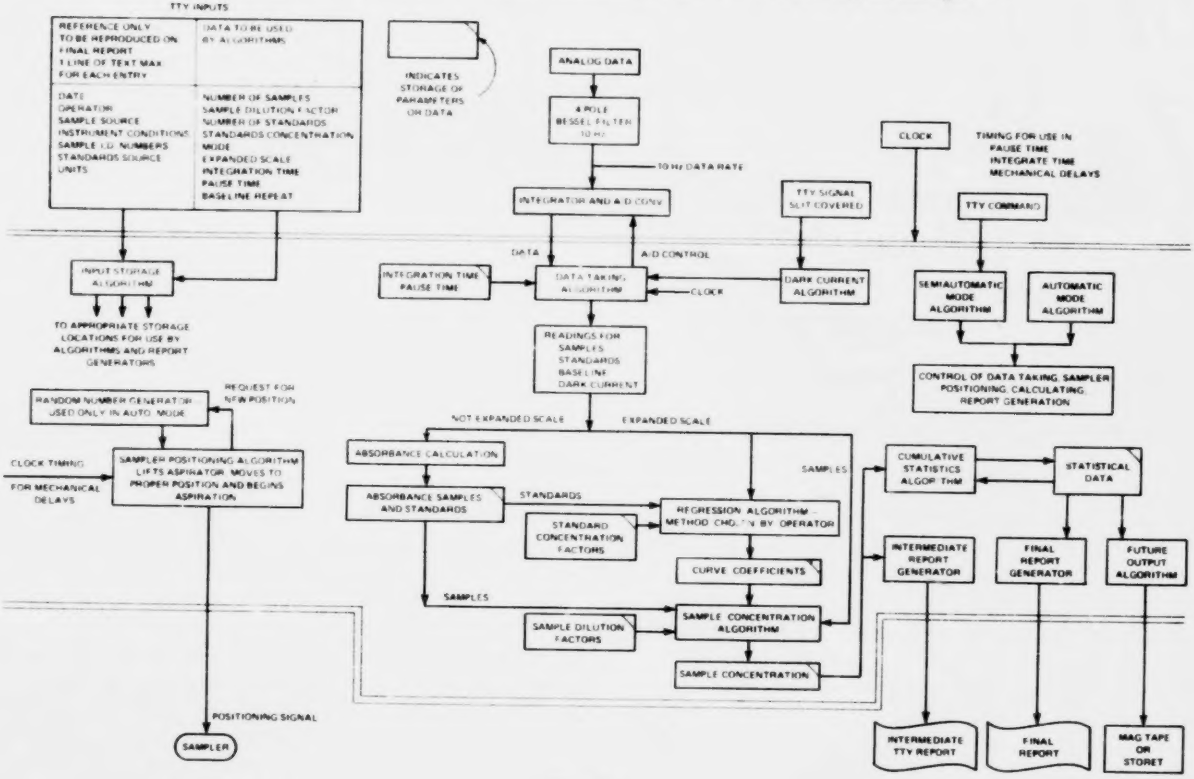
X = Present position (stored).

Y = Next position - calculated from teletype input or random number generator.

Z = Total number of positions.

40/Z = W = Number of steps per position. Wheel positions numbered zero to Z - 1.

Figure 1
Atomic Absorption Instruments Control Algorithm For Positioning the Sampler



TTY INPUTS
REFERENCE ONLY TO BE REPRODUCED ON FINAL REPORT 1 LINE OF TEXT MAX FOR EACH ENTRY
DATA TO BE USED BY ALGORITHMS
DATE
OPERATOR
SAMPLE SOURCE
INSTRUMENT CONDITIONS
SAMPLE I.D. NUMBERS
STANDARDS SOURCE
UNITS
NUMBER OF SAMPLES
SAMPLE DILUTION FACTOR
NUMBER OF STANDARDS
STANDARDS CONCENTRATION
MODE
EXPANDED SCALE
INTEGRATION TIME
PAUSE TIME
BASELINE REPEAT
INDICATES STORAGE OF PARAMETERS OR DATA
ANALOG DATA
4 POLE BESSEL FILTER 10 Hz
10 Hz DATA RATE
INTEGRATOR AND A/D CONV
CLOCK
TIMING FOR USE IN PAUSE TIME INTEGRATE TIME MECHANICAL DELAYS
TTY SIGNAL SLIT COVERED
TTY COMMAND
INPUT STORAGE ALGORITHM
DATA
A/D CONTROL
INTEGRATION TIME PAUSE TIME
DATA TAKING ALGORITHM
CLOCK
DARK CURRENT ALGORITHM
SEMIAUTOMATIC MODE ALGORITHM
AUTOMATIC MODE ALGORITHM
TO APPROPRIATE STORAGE LOCATIONS FOR USE BY ALGORITHMS AND REPORT GENERATORS
READINGS FOR SAMPLES STANDARDS BASELINE DARK CURRENT
CONTROL OF DATA TAKING, SAMPLER POSITIONING, CALCULATING, REPORT GENERATION
RANDOM NUMBER GENERATOR USED ONLY IN AUTO. MODE
REQUEST FOR NEW POSITION
NOT EXPANDED SCALE
EXPANDED SCALE
CLOCK TIMING
FOR MECHANICAL DELAYS
SAMPLER POSITIONING ALGORITHM LIFTS ASPIRATOR, MOVES TO PROPER POSITION AND BEGINS ASPIRATION
ABSORBANCE CALCULATION
SAMPLES
CUMULATIVE STATISTICS ALGORITHM
STATISTICAL DATA
ABSORBANCE SAMPLES AND STANDARDS
STANDARDS
REGRESSION ALGORITHM - METHOD CHOSEN BY OPERATOR
STANDARD CONCENTRATION FACTORS
CURVE COEFFICIENTS
INTERMEDIATE REPORT GENERATOR
FINAL REPORT GENERATOR
FUTURE OUTPUT ALGORITHM
SAMPLES
SAMPLE DILUTION FACTORS
SAMPLE CONCENTRATION ALGORITHM
SAMPLE CONCENTRATION
POSITIONING SIGNAL
SAMPLER
INTERMEDIATE TTY REPORT
FINAL REPORT
MAG TAPE OR STORET

Figure 2
Functional Schematic for Auto Analyzers

SOFTWARE COMPATIBILITY IN MINICOMPUTER SYSTEMS

By John O.B. Greaves

INTRODUCTION

Suppose that we wanted to build a Behavior Studying Machine to observe and quantize the motions and responses of various aquatic organisms. We would also like to direct the machine to record the images of some data, to display the images, and to manipulate these pictures into some usable form. Then we would like to be able to attach some statistical significance to the results. Once the machine is built, the method of its construction may be deemed irrelevant until some other person requests one like it. Or, if the machine breaks, as machines do, then the method of construction plays a significant role in how much time it takes to repair it and what training is required to do so.

To communicate our wishes and commands to the machine, we develop a language called the Behaviorial Response Language or BRL. Since we wish to use the machine in an interactive fashion with graphics capabilities, the language should be interpretive rather than translative. This language will be executed on the Behavior Studying Machine. In a purely top-down approach to the design of the machine, this language would be the starting point. We could define layers of metalanguages until, at the bottom, there would be some mechanical and/or electronic hardware to execute the commands from higher levels. With a purely bottom-up approach, we could choose a logic line of integrated circuits or relays and build upon that a nanoprograming language, a microprograming language, a conventional machine language, and on up until the BRL could be executed.

BUILDING A VIRTUAL MACHINE - THE BEHAVIOR STUDYING MACHINE

In building a realizable system, the systems architect often employs a middle-out design, with good reason. Being knowledgeable about the requirements of the top and the possibilities at the bottom, the system can grow in both directions through the levels of structure toward a realizable system. We set out to build the Behavior Studying Machine without knowing what computer would eventually perform the lower level operations. Because of the high data rates for storing images, a memory and a disk had to be local. Because of the cost, the computer had to be mini/midi. The FORTRAN IV language was chosen as the primary metalanguage because reasonably standardized translators exist on many machines; for example, there are several "virtual FORTRAN" machines on the market. For another reason, FORTRAN IV is a translative rather than an interpretive language and gives us a run time speed advantage. While our language, the Behavioral Response Language, is interpretive, the modules it executes are translated from FORTRAN to a lower language prior to run time and, therefore, they need not be scanned for lexical and syntactical errors. To achieve ultimate speed advantage, we could provide a translator to translate our metalanguage directly into microcode for the host computer, but perhaps the time payoff would be long in coming.

To assist in describing the Behavior Studying Machine, we will provide the following top-down sketch. First, the keyboard function names include INPUT <parameters>, PLOT < >, FIND PATHS < >, SQUARE < >, EDIT < >, SUM < >, and LIV (for Linear Velocity). These modules are coded as FORTRAN subroutines and provide the basic elements of the BSM. Thus, when new functions are added to the system, they may be called by name and may receive their keyboard arguments in a labeled common block that has been scanned, decoded, and placed into fixed alphabetic and numeric fields. These modules, in turn, interface with the next lower level via subroutine calls to the file handlers. The subroutines are described as follows: BCREAT to create a new file, BOPEN to open (or try) an existing file, BCLOSE to close a file, GETV to get a vector (a variable length record) from a file, and WRITEV to write a vector into a file. At this level, if the programer only abides by the rules of the interface, the entire business of managing the buffer space becomes transparent. Thus, all existing and proposed keyboard functions are machine-independent. The file handlers are therefore conceived as the next level down, but they, too, are coded in FORTRAN. The file handlers make calls upon the next lower level in two forms, the buffer managers and the operating system file subroutines. The first of these is made up of buffer manager modules. They perform the virtual memory-like function of swapping in and out sections of the disk-based file data as requested. Typical **calls** to the buffer manager modules are SEARCH to search for a vector, HILOV to determine the highest and lowest vector resident in the buffer, and GETBUF & RELBUF to get and to release buffer space respectively. Typical operating system calls are OPEN, CLOSE, RDBLK and WRBLK to read and write blocks on the disk. The buffer manager routines are also coded in FORTRAN. They allocate and deallocate another labeled common block dimensioned at about eight thousand words to allow that many integers or half as many real data elements resident at any given time. They will execute these routines on any "virtual FORTRAN" machine. Calls to the operating system file handlers are both machine-dependent and operating-system-dependent. These sections must be suitably interfaced to the operating system of the host computer, which we found to be a nonmonumental task. The primitive operations are the same; for example, open, close, read, and write.

At the lowest level, assembly language programing, we have two classes of subprograms. The first class involves augmenting the FORTRAN language with three primitive subprograms: (1) the logical LSTEQ operator to compare two N character strings to determine if they are the same (.TRUE.) or different (.FALSE.), (2) UNPACK to unpack two bytes into two integer words, and (3) PACK to pack two bytes into one integer word. All three subprograms were also coded in FORTRAN, but the assembler versions were used for aesthetic reasons and for the checkout of the FORTRAN-assembler interface. The second class of assembler subprograms exist as software drivers for the special-purpose video to digital converter, the "Bugwatcher." The Bugwatcher is connected to the computer via a direct memory access (DMA) channel, for speed. These subprograms have well-defined machine-independent interfaces. They are: BWSEND to send a command word to the Bugwatcher to set the input frame rate or the video threshold, to reset the hardware, or initiate or terminate the transfer of data from the Bugwatcher; BWSETS to set up for a Single buffering operation; BWSETD to set up for a Double buffering operation; BWONS to turn on the DMA channel for a Single buffer; BWOND to turn on the channel for Double buffering; BWWAIT to wait for the buffer to fill and interrupt the processor; and BWOFF to disable the interruption and terminate the DMA transfer. The simpler single buffering programs are used for initial checkout and for the LIVE keyboard function to check that all hardware-software systems are working properly. The double buffering subprograms are used to transfer data from the DMA input to disk for later retrieval and analysis.

THE LOWER LEVELS

We are currently implementing the Behavioral Response Language on three separate virtual machines: the DOS-9 operating system on a PDP 11/45, the RDOS operating system on a Data General ECLIPSE S/200 and, most recently, the RT-11 operating system on the same PDP 11/45. To bring the software up on the ECLIPSE RDOS operating system is roughly a one to two mythical man-months (MMM) effort with another month for hardware connections cabling, interfacing, and checkout. The time estimated to change operating systems on the DEC PDP 11/45 is one MMM, which seems reasonable with the progress seen to date.

CONCLUSION

One design criterion has been to bridge the gap between the interpretive Behavioral Response Language and the specially designed hardware, the Bugwatcher, using software as portable as possible. To accomplish this, we developed a structure of layered software, each with a well-defined interface that could be implemented with relative ease on most minicomputers. Software, being pliable only in its formulative stages, can be shaped for this purpose. A more complete description of the prototype "Bugsystem" can be found in the following references list.[1 2]

REFERENCES

1 Davenport, D., Culler, G.J., Greaves, J.O.B., Foreward, R.B. and Hand, W.G., "The Investigation of the Behavior of Microorganisms by Computerized Television," *IEEE Trans. Biomed. Eng.*, Vol. BME-17, July 1970, pp. 230-237.

2 Greaves, John O.B., "The Bugsystem: The Software Structure for the Reduction of Quantized Video Data of Moving Organisms," *Proc. of the IEE*, Vol. 63, No. 10, October 1975.

SUMMARY OF DISCUSSION PERIOD - PANEL I

The question and answer session was brief as a result of the extended paper presentation session. A summary of the discussion is presented below.

Mini- and Microcomputers Versus Large Computer Systems

When asked whether minicomputers and microcomputers would eliminate the need for large-scale data base systems, the panel felt strongly that they would not. With the introduction of the low cost microprocessor, even the lowest cost instruments in the laboratory, including the pH meter, will inevitably be automated. The increased automation will produce machine-readable information. There will be an increased requirement to transfer the information from facility to facility and to store the data in a large data base for subsequent retrieval and analysis.

Laboratory Automation Serves The Analyst

Spontaneous laughter followed when the panel was asked if laboratory automation systems were for the benefit of the "Boss" to keep track of what his people were doing at the bench and were just a waste of the analyst's time. The responses were varied, but all contained the common theme that all of the laboratory automation systems observed in operation by the panel members served the analyst. The analyst was assisted in data reduction and in data handling, and quality assurance increased with laboratory automation systems.

Training

The panel members were asked about the difficulty in training personnel in utilizing laboratory automation systems that they had implemented. The members agreed that the training process involved few difficulties. Generally, the personnel utilizing laboratory automation systems were briefed on the flow of information from the instrument to the printed output. This indoctrination was followed with on-the-job training with satisfying results. In designing one system, the user helped define what interaction would be necessary for optimum system/user interaction. Another system contains online assistance in the form of "Help" commands which the user may invoke at any time.

Computers for Research

The question that evoked the most discussion was whether ORD was becoming a computer-oriented organization rather than an environmental research organization. The panel response was an emphatic "No." It is very difficult for EPA to complete its mission without getting more heavily involved in utilizing laboratory automation systems because of current manpower limitations. The computer will be used as a tool just as the atomic absorption spectrometer is commonly used; ORD is not being accused of being an "instrument house" because it uses atomic absorption instruments. According to Frazier of the Lawrence Livermore Laboratories, EPA is the leader for the country in laboratory automation systems and the high cost of the equipment it has and the systems it is implementing. The techniques ORD is implementing are forerunners of those that even municipalities will be using as the proliferation of microprocessors continues. It was suggested that the systems will be available for 10 to 20 percent of today's developmental cost.

Laboratory Feasibility Studies

Given that feasibility studies consume considerable resources, the panel was asked when the laboratory manager performs a feasibility study. The panel response to this very important question was less than adequate, mainly because there is no logical set of rules to follow to determine when a feasibility study is to be performed. A study could be included as part of the ongoing effort in the development of project plans which are formed to meet EPA's mission, or the manager may find that there is no way to complete his project without the aid of a laboratory automation system.

Laboratory Instrumentation

The question was raised as to which laboratory instruments lend themselves to automation with microprocessors. It was the opinion of the panel that with the advent of the microprocessor, no instrument in the laboratory would be excluded from automation.

Standardized Laboratory Automation System

The final question inquired of the panel whether EPA should have a standardized laboratory automation system. It seemed inconceivable to the panel members that there would be a standard automation system since each of the laboratories has such diverse research objectives. The computer industry is changing and the standardization of hardware, software, and especially hardware interfaces may be realized in the future, but EPA, as an enforcement agency, cannot afford to wait and standardize.

LABORATORY DATA MANAGEMENT

By William L. Budde

Data management is a term that has many different meanings. To the accountant, it means payroll and inventory control. In the environmental field, data management could mean archival storage of environmental data, trend analyses and statistics, environmental quality indexes, mathematical modeling, or a host of other activities.

For purposes of this discussion, data management has a limited meaning. We will be concerned with the management of information related to the operations of a laboratory that makes chemical, physical, and biological measurements on environmental samples. In other words, our concern is limited to the computerized handling of information about samples that are currently in process in a laboratory.

Laboratory data management (LDM) could start several months before sampling begins when a project manager defines sampling sites, sampling dates or times, and analytes to be measured. Subsequent entries may include project/sample numbers, maximum allowable holding times, values from field measurements, and a variety of other information. At any time, management may require a laboratory workload projection based on all current and expected samples. After samples are received at the laboratory, daily work lists may be generated for specific analytes with references to potential interferences; for example, a particularly high trace metal concentration could be noted on the nutrients work list. As measurements are completed, data must be sorted by project, interim status reports generated, and final reports printed. At periodic intervals, management may require other reports to summarize quality control or allocate operations costs.

A traditional approach to LDM is manual entry of all information into a computerized data base located at a large data processing center. Manual entry includes automatic but offline transfers, such as punched card input, punched paper tape input, magnetic mark/print sensing, and keyboard entry from a time sharing computer terminal. One of the principal advantages of this mode of operation is that it permits the utilization of large computer systems with massive memories and a variety of peripheral equipment, including high-speed printers, plotters, and microfilm/microfiche output. In addition, in recent years, general purpose data base management software has become available for use on medium- to large-scale computer systems. The disadvantages of this approach include the potentially significant error rate associated with manual data entry systems and the relatively slow turnaround times that are standard on many batch computing systems. Of course, careful verification of manually entered data and fast batch turnarounds are possible, but usually at a significantly higher cost. Several other panelists have described their operations and experiences with the traditional approach to LDM.

An alternative to the traditional approach to LDM is local data management with a minicomputer. This is a viable choice because of the development of relatively inexpensive minicomputers for data acquisition, data reduction, and control of analytical instruments.

Instrument automation permits a substantial improvement in quality assurance by transferring instrument output signals to the computer via direct electrical connections. The signals are processed in digital form to generate measured values. These operations eliminate errors from hand measurements of peak heights or areas, hand or desk calculator computations, manual transcriptions, and coding in computer readable form. The efficiency of the whole process permits the incorporation of many quality assurance checks as an integral part of the instrument operating system. Typical quality checks include frequent analyses of check standards, replicates, spikes, blanks, and reagent blanks. These data may be assessed rapidly by statistical methods and the accuracy and precision compared with established accuracy and precision data for that method, instrument, and operator.

It is very attractive to develop an LDM system using a computer in the laboratory that has direct access to the online acquired data. This approach has the advantage of preserving the quality of the data since no manual transfers would be required. Another advantage is that fast turnaround times would be possible at relatively low cost. Few, if any, LDM systems have been developed that coexist with laboratory instrument automation systems. However, a system of this type is under development by EPA and several panelists at this session have discussed progress to date.

There are three principal options for local LDM: (1) background processing during the day in the laboratory automation computer, (2) overnight processing in the laboratory automation computer, and (3) the employment of a second minicomputer in the laboratory for data management. The second processor would have access to the laboratory data via a shared mass storage device or a high-speed data channel. Figure 1 shows a proposed laboratory computer network for instrument automation, LDM, and subsequent transfer of data to a large computer for non-LDM operations. The large computer system is a traditional data processing center.

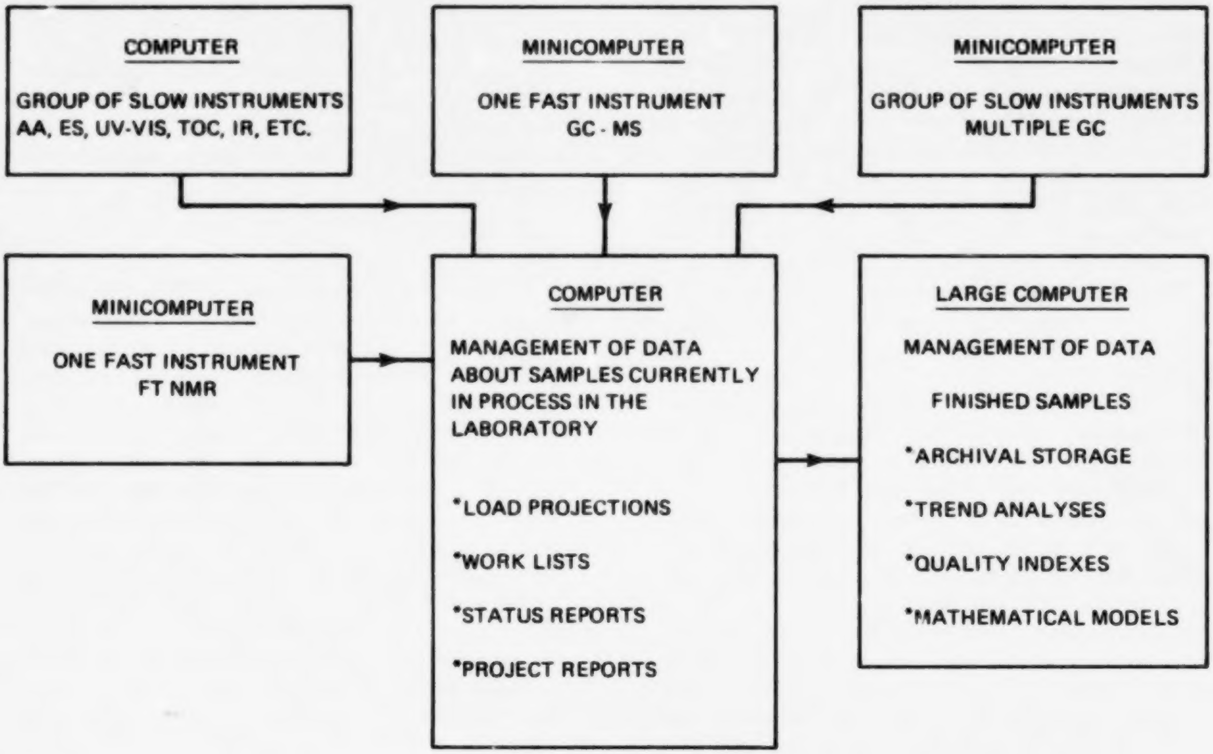

COMPUTER
GROUP OF SLOW INSTRUMENTS
AA, ES, UV-VIS, TOC, IR, ETC.
MINICOMPUTER
ONE FAST INSTRUMENT
GC - MS
MINICOMPUTER
GROUP OF SLOW INSTRUMENTS
MULTIPLE GC
MINICOMPUTER
ONE FAST INSTRUMENT
FT NMR
COMPUTER
MANAGEMENT OF DATA ABOUT SAMPLES CURRENTLY IN PROCESS IN THE LABORATORY
*LOAD PROJECTIONS
*WORK LISTS
*STATUS REPORTS
*PROJECT REPORTS
LARGE COMPUTER
MANAGEMENT OF DATA
FINISHED SAMPLES
*ARCHIVAL STORAGE
*TREND ANALYSES
*QUALITY INDEXES
*MATHEMATICAL MODELS

Figure 1
Laboratory Computer Network

SUSPENDED PARTICULATE FILTER BANK AND SAMPLE TRACKING SYSTEM

By Thomas C. Lawless

INTRODUCTION

The National Air Surveillance Network (NASN) was established in 1953 by the Public Health Service in cooperation with State and local health departments. Today there are approximately 300 urban and nonurban sampling stations operating as part of the NASN. Part of the network monitoring activity is devoted to sampling total suspended particulate matter. These samples are obtained with a high volume sampler which draws air through an 8-by 10-inch glass-fiber filter. The sampling schedule used by the network calls for the collection of one 24-hour sample each 12 days, or approximately 30 samples a year. The Filter Bank System was developed to act as an inventory system for monitoring the status of samples during and after analysis, i.e., storage.

FILTER BANK SYSTEM

Exposed filters from each sampling site are sent to its respective regional office, which, in turn, forwards them to the Environmental Monitoring and Support Laboratory at the Environmental Research Center in North Carolina on a monthly or quarterly basis. Attached to each filter folder is an Air Quality Data Bank form for particulate data (Figure 1). This form contains the filter number, the sampling date, the air volume, and the 12-digit SAROAD (Storage and Retrieval of Aerometric Data) station code depicting the exact location of the sampling site. SAROAD is a data storage system which is part of the Aerometric and Emissions Reporting System (AEROS). The Filter Bank System requires that all filters be validated, e.g., checked for tears and so forth, and that all filter cards be checked for completeness of sampling information. Subsequently, these filters are entered into the Filter Bank.

The Filter Bank is a data processing system developed using SYSTEM 2000* on the Univac 1110. It uses the Immediate Access, the Report Writer, and the Program Language Interface aspects of SYSTEM 2000. Use of SYSTEM 2000 allows specification without restriction of elements in the data base which are key fields and identification of hierarchical relationships among elements in the data base. Information retrieval is dependent upon components which are declared key fields. Data security is maintained by password control to the data base and additional password control to each component. The Filter Bank System is centered around a data base designed to use the standard SAROAD coding procedures. There is a logical entry (tree structure) for each site in the NASN. As shown in Figure 2, data sets, which are subordinate to each site, contain the filter information. Associated with each of these data sets is another group of data sets containing the pollutant, method, and units information and its specific analytical result.

DATA ENTRY

Filter information is entered into the Filter Bank through a interactive terminal by using a prompting procedure. This procedure contains a SYSTEM 2000 FORTRAN interface program which allows data entry by non-ADP personnel. It also provides the editing capabilities for the Filter Bank System. The site code, sampling date, filter number, filter type, and air volume are entered. Before the next entry is possible, the Filter Bank System checks for a valid site code, a valid date, and for completeness of entry. The entry is rejected if these requirements are not met, and a short description of the error is displayed on the terminal. The entry is also rejected if the site/date combination is not unique to the Filter Bank. If the error is obvious, the correct information can be reentered immediately and when accepted, the system replies by printing the 4-letter code it is assigning to this particular filter. Use of a 4-letter coding scheme can accommodate over 400,000 filters. (The system is easily modified to a 5-letter code which will accommodate over 10 million filters.)

The update session is terminated by executing another SYSTEM 2000 procedure. The first part of this procedure is a program which produces an update log listing the site and date of each accepted filter. The program also produces a label to be attached to each filter folder, and this label contains the information on the data card together with the unique 4-letter code. Subsequently, labeled filters are stored until analysis. The second part of this procedure is a SYSTEM 2000 Immediate Access program which backs up the Filter Bank onto a magnetic tape to protect against loss due to system or machine failure. This precaution was developed out of necessity during the early days of the Univac.

---

* SYSTEM 2000 - MRI Systems Corporation.

SAMPLE ANALYSIS

For the analysis of nitrates, sulfates, and ammonia, a portion of each filter is cut and sent to the laboratory accompanied by its assigned unique 4-letter code. There, sets of samples are arranged for the Technicon Auto Analyzer so that there are two complete sets of standards, a standard after every tenth sample, and a series of blanks. Periodically, quality audit samples (previously analyzed samples) are included. Upon completion of the analyses, the results of each set of samples, standards, and blanks, in micrograms per milliliter, with their associated filter codes, are transcribed onto a specially designed form. In the event of dilution of a particular sample, the dilution factor is coded. Likewise, if a color analysis was performed, this result is also coded. Then the Filter Bank System accepts these entries, separates them into data samples and quality assurance audit samples, and processes them accordingly. The system produces a listing of the data samples with final concentrations expressed in micrograms per cubic meter, calculated by using the previously stored air volumes. A statistical summary of the standards, blanks, and audit samples accompanies each tabulation. These listings are returned to the laboratory for validation and determination of acceptability of the data (Figure 3). Using this summary and information from previous standard analyses, an analytical data quality indicator is presently being developed which will quantitatively determine the acceptability of the analysis.

For the analysis of metals, using the Optical Emissions Spectrograph, quarterly composite samples are prepared using information supplied by the Filter Bank System. A SYSTEM 2000 FORTRAN program prepares this information on site quarters which have met established criteria regarding number and spacing of filters. The filters which make up the valid sample and the average air volume for the composite are listed. The program also sets the composite flag of each sample in the composite for future reference. Upon completion of the laboratory analysis, a magnetic tape containing the results is passed on to the Filter Bank System. The tape is processed and a data tab is produced showing duplicate analytical results, the average result, and the percent relative standard deviation for all metals of each site quarter.

The Filter Bank System then stores the date of analyses and the sample results (nitrates, sulfates, ammonia, and metals) using the standard SAROAD formats for pollutant, method, and units with the other filter information (Figure 4). In addition to storing all the information in the Filter Bank, the system produces a file of SAROAD formatted records to be passed on to the National Aerometric Data Bank.

REPORTING

Periodically, computer printouts are produced to inform the laboratory of filters or specific analyses that were inadvertently overlooked or lost in processing. This report also includes quarterly composite samples which became valid and eligible for analysis due to late arriving filters. This safeguard assures that all required analyses are performed on all filters received. The Filter Bank System inventories the data base and lists, by site, all dates for which filters have been received, those filters which have been analyzed, and those pollutants which have been analyzed for, as well as analytical results with yearly averages.

Various information retrieval techniques are possible. The Immediate Access feature of SYSTEM 2000 provides a user-oriented language with which a nonprogramer may express his request for Filter Bank information, This feature is highly suited for interactive use from a remote keyboard. SYSTEM 2000 allows access to the data base by multiple users simultaneously. Using this feature, one can determine the status of any filter, the analyses which have already been done, and the results of these analyses.

The following examples demonstrate the usefulness of the Immediate Access feature. The response time for each of these examples varies with machine load but averages less than 10 seconds. Component numbers correspond to those in Figure 2.

(1) To print the filter information for which either the site code and date are known or the 4-letter filter code is known:

```
PRINT SAMPLE WHERE SITE EQ
056980004A01 AND INDATE EQ 12/25/74:

210*  HIV
220*  7
230*  12/25/1974
240*  0
250*  1044964
260*  AIVR
270*  3063
280*  1
```

```
PRINT SITE, STATE, CITY, SAMPLE
WHERE FILCOD EQ AIVR:

1*      056980004A01
2*      CALIFORNIA
3*      SAN JOSE

210*    HIV
220*    7
230*    12/25/1974
240*    0
250*    1044964
260*    AIVR
270*    3063
280*    1
```

(2) To print the pollutant information for a site/ date combination.

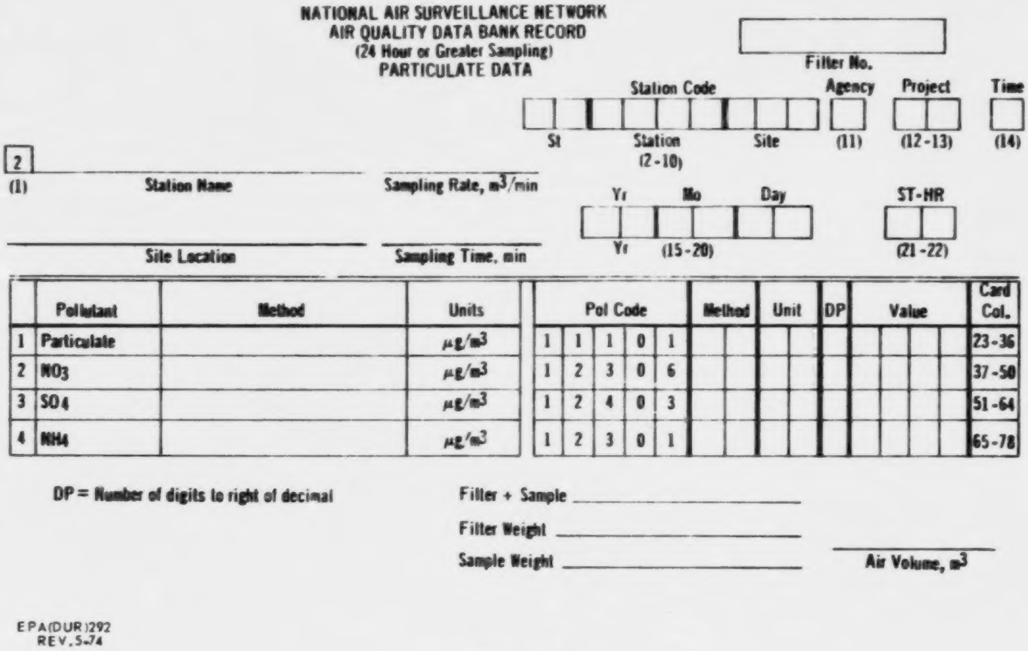
```
PRINT POLUT WHERE SITE EQ
056980004A01 AND INDATE EQ 12/25/74:

310*    12306
320*    92
330*    1
340*    9.1500
350*    2
360*    04/29/1975

310*    12403
320*    91
330*    1
340*    3.0000
350*    1
360*    04/29/1975

310*    12301
320*    92
330*    1
340*    .6900
350*    2
360*    04/29/1975
```

(3) To print the date of filters which have not been analyzed for a site:

```
PRINT INDATE WHERE SITE EQ
056980004A01 AND VALUE FAILS:

230*     01/06/1975
230*     01/18/1975
230*     01/30/1975
230*     02/11/1975
230*     02/23/1975
230*     03/07/1975
230*     03/19/1975
230*     03/31/1975
230*     04/12/1975
230*     04/24/1975
230*     05/06/1975
230*     05/18/1975
230*     05/30/1975
230*     06/11/1975
230*     06/23/1975
230*     07/05/1975
230*     07/17/1975
230*     07/29/1975
230*     08/10/1975
230*     08/22/1975
230*     09/03/1975
230*     09/15/1975
230*     09/27/1975
```

(4) To print the dates and air volumes for a site:

```
PRINT INDATE-FILCOD WHERE 01 EQ
056980004A01 AND INDATE LT 01/01/75:

230*     12/06/1973
260*     ACGC
230*     12/18/1973
260*     ACGD
230*     12/30/1973
260*     ACGE
230*     01/11/1974
260*     ACGF
230*     01/23/1974
260*     ACG
230*     02/04/1974
260*     ADXB
230*     02/28/1974
260*     ADXC
230*     03/12/1974
260*     ADXD
230*     03/24/1974
260*     ADXE
230*     04/17/1974
260*     ADXF
230*     04/29/1974
260*     ADXG
230*     05/11/1974
260*     ADXH
230*     05/23/1974
260*     ADXI
```



NATIONAL AIR SURVEILLANCE NETWORK
AIR QUALITY DATA BANK RECORD
(24 Hour or Greater Sampling)
PARTICULATE DATA

Filter No.

Station Code: St, Station (2-10), Site
Agency (11)
Project (12-13)
Time (14)

2 (1) Station Name
Sampling Rate, m^3/min

Yr Mo Day (15-20)
ST-HR (21-22)

Site Location
Sampling Time, min

	Pollutant	Method	Units	Pol Code	Method	Unit	DP	Value	Card Col.
1	Particulate		$\mu g/m^3$	1 1 1 0 1					23-36
2	NO_3		$\mu g/m^3$	1 2 3 0 6					37-50
3	SO_4		$\mu g/m^3$	1 2 4 0 3					51-64
4	NH_4		$\mu g/m^3$	1 2 3 0 1					65-78

DP = Number of digits to right of decimal

Filter + Sample ____________________
Filter Weight ____________________
Sample Weight ____________________
Air Volume, m^3

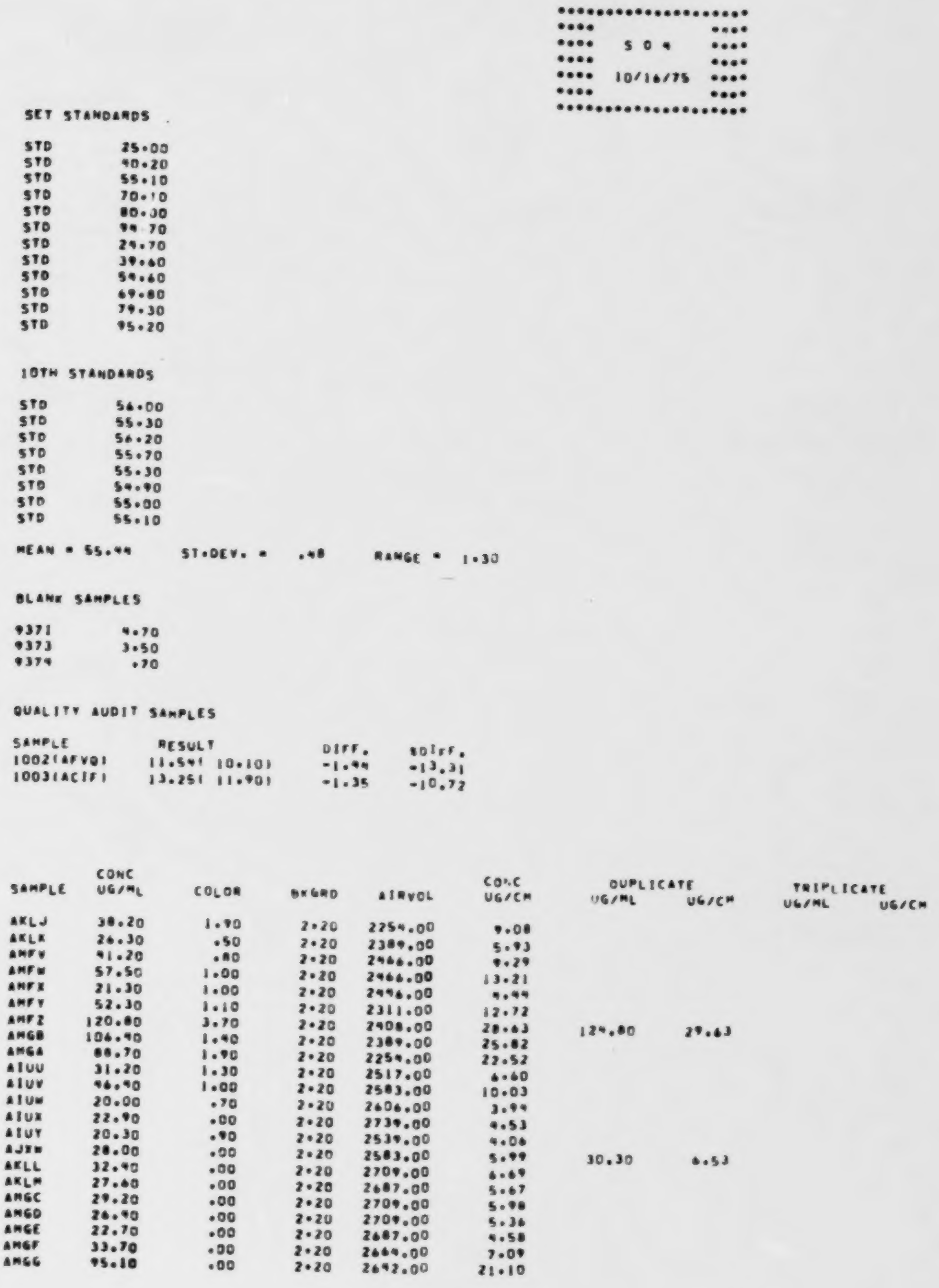
EPA(DUR)292
REV.5-74

Figure 1
Air Quality Data Bank Form

DATA BASE

- 1* SITE (Key)
- 2* STATE
- 3* CITY
- 20* SAMPLE (Repeating Group)
  - 210* FILTER TYPE (Key)
  - 220* SAMPLE INTERVAL
  - 230* SAMPLE DATE (Key)
  - 240* START HOUR
  - 250* FILTER NUMBER
  - 260* FILTER CODE (Key)
  - 270* AIR VOLUME
  - 280* COMPST (Key)
  - 30* POLLUTANT (Repeating Group in Sample)
    - 310* POLLUTANT CODE (Key)
    - 320* METHOD OF ANALYSIS
    - 330* UNITS
    - 340* VALUE (Key)
    - 350* NO. DECIMAL PLACES
    - 360* DATE OF ANALYSIS

Figure 2
Data Base Structure



```
                                                        *******************
                                                        ****           ****
                                                        ****   S O 4   ****
                                                        ****           ****
                                                        ****  10/16/75 ****
                                                        ****           ****
                                                        *******************
SET STANDARDS

STD          25.00
STD          40.20
STD          55.10
STD          70.10
STD          80.00
STD          94.70
STD          24.70
STD          39.60
STD          54.60
STD          69.80
STD          79.30
STD          95.20

10TH STANDARDS

STD          56.00
STD          55.30
STD          56.20
STD          55.70
STD          55.30
STD          54.90
STD          55.00
STD          55.10

MEAN = 55.44     ST.DEV. =    .48     RANGE =  1.30

BLANK SAMPLES

9371          4.70
9373          3.50
9374           .70

QUALITY AUDIT SAMPLES

SAMPLE          RESULT            DIFF.     %DIFF.
1002(AFVQ)    11.54( 10.10)      -1.44     -13.31
1003(ACIF)    13.25( 11.90)      -1.35     -10.72


         CONC                                          CONC
SAMPLE   UG/ML      COLOR     BKGRD     AIRVOL        UG/CM      DUPLICATE            TRIPLICATE
                                                              UG/ML     UG/CM      UG/ML     UG/CM

AKLJ      38.20      1.90      2.20     2254.00       9.08
AKLK      26.30       .50      2.20     2389.00       5.93
AMFV      41.20       .80      2.20     2466.00       9.29
AMFW      57.50      1.00      2.20     2466.00      13.21
AMFX      21.30      1.00      2.20     2496.00       4.44
AMFY      52.30      1.10      2.20     2311.00      12.72
AMFZ     120.80      3.70      2.20     2408.00      28.63    124.80     29.63
AMGB     106.90      1.40      2.20     2389.00      25.82
AMGA      88.70      1.90      2.20     2254.00      22.52
AIUU      31.20      1.30      2.20     2517.00       6.60
AIUV      46.90      1.00      2.20     2583.00      10.03
AIUW      20.00       .70      2.20     2606.00       3.94
AIUX      22.90       .00      2.20     2739.00       4.53
AIUY      20.30       .90      2.20     2539.00       4.06
AJXW      28.00       .00      2.20     2583.00       5.99     30.30      6.53
AKLL      32.90       .00      2.20     2709.00       6.69
AKLM      27.60       .00      2.20     2687.00       5.67
AMGC      29.20       .00      2.20     2709.00       5.98
AMGD      26.90       .00      2.20     2709.00       5.36
AMGE      22.70       .00      2.20     2687.00       4.58
AMGF      33.70       .00      2.20     2664.00       7.09
AMGG      95.10       .00      2.20     2692.00      21.10
```

Figure 3
Data Listing

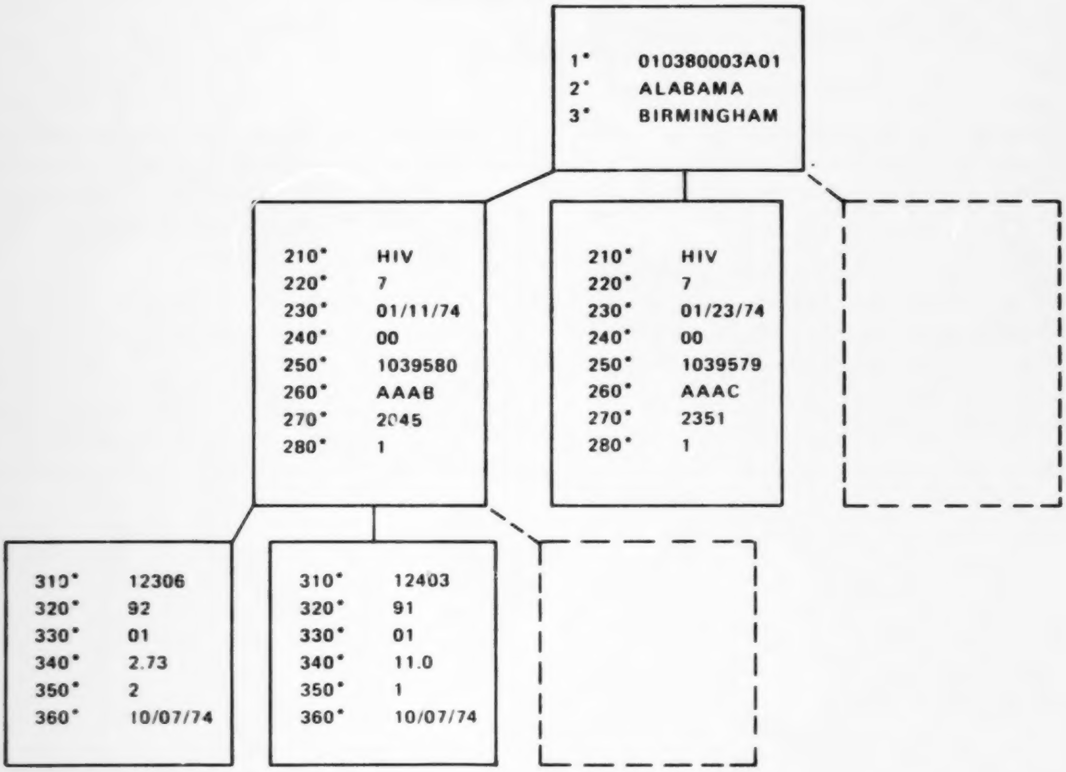

1* 010380003A01
2* ALABAMA
3* BIRMINGHAM
210* HIV
220* 7
230* 01/11/74
240* 00
250* 1039580
260* AAAB
270* 2045
280* 1
210* HIV
220* 7
230* 01/23/74
240* 00
250* 1039579
260* AAAC
270* 2351
280* 1
310* 12306
320* 92
330* 01
340* 2.73
350* 2
360* 10/07/74
310* 12403
320* 91
330* 01
340* 11.0
350* 1
360* 10/07/74

Figure 4
Logical Entry

EIGHT YEARS OF EXPERIENCE WITH AN OFF-LINE LABORATORY DATA MANAGEMENT SYSTEM USING THE CDC 3300 AT OREGON STATE UNIVERSITY

By D. Krawczyk

ADP is the acronym for automatic data processing.[1] ADP use in chemistry laboratories indicates considerable labor savings through its proper use. At the last ORD ADP workshop, Byram et al. reported on one facet of ADP: the combination of an automated colorimetric system with a computer.[2] In addition, this writer discussed SHAVES, a broad aspect of data management.[3] Whether printouts, paper tapes, or published reports are used to report data, the important products from an analytical laboratory are valid results. Let me quote from Rudyard Kipling:

> "The careful text-book measure
> (Let all who build beware!)
> The load, the shock, the pressure
> Material can bear.
> So when the buckled girder
> Lets down the grinding span,
> The blame of loss or murder,
> Is laid upon the man
> Not on the stuff–the Man"[4]

As Kipling points out in the verse, blame for a fallen bridge cannot be placed on education, loading, shock, or other known variables but on the builder-designer. The number of labor saving devices used in the laboratory is unimportant. If the answer is invalid, then man bears the responsibility. Eight years of experience at the Corvallis laboratory have resulted in some novel approaches to handling data, especially in quality assurance aspects.

The availability of the Oregon State University computer, literally across the street, provided the laboratory group with an experimental tool to aid in handling data. Initially, ADP use followed the sample control and verification principle reported by Krawczyk et al.[5] As the demand for flexibility, responsiveness, and personnel limitations increased, greater emphasis was placed on the use of ADP as a tool. Getting the job done for the least cost resulted in using automated chemical analytical systems connected with ADP. During the 8 years of our experience with ADP, the need arose each month for changing or modifying one or another subprogram to improve quality assurance of data. Not only has ADP permitted the laboratory to respond more effectively but also has made the job easier. The first and foremost purpose of our use of ADP is the production of quality data. From collection of samples to final reporting of data, quality assurance is a way of operation within the Corvallis laboratory.

The 8 years of experience began modestly with approximately 6,000 samples and 40,000 analyses from six to seven projects per year. Graphical representations of projects, samples, and results since 1972 are shown in Figures 1, 2, and 3. What was accomplished 8 years ago in 6 months was exceeded in 1 month in calendar 1973, 1974, and 1975. In 1972, the workload was performed with a staff of 17 permanent employees. In calendar 1975, the staff was reduced to 12 permanent people with increased use of temporary employees to accomplish specific short-term tasks. The shift from permanent to temporaries required further refinements in use of ADP, especially from the quality assurance aspect.

An illustration of a change made to provide more rapid response with quality is a change made in the scheduling subroutine. During the last 3 years, the computer was used to schedule the analysis of samples for forms of nitrogen and phosphorus through our automated subroutine. Initially, the computer produced a run with samples scheduled for analysis ordered as the samples were brought into the laboratory. Thus, fresh waters were mixed with marine waters; waste-water treatment plant effluent samples preceded and followed base water used in bioassay experiments. From a computerized standpoint, first in, first out, was good management; however, the technical problems generated by such an approach were difficult to handle. Very quickly the scheduling was modified to permit choice of samples by project designation. Project assignment was usually made based on type of sample. The analyst chose his mixture of samples and processed similar types at one time. The procedure of first in, first out, was then changed to permit combining similar samples into a production run.

In the examples of outputs of quality assurance in Figure 4, a listing of intercomparison errors is presented for the week ending October 8, 1975. During this period, approximately 3,000 results were processed through program "TECKNICON." Approximately 1,300 results were reported by analysts for metals, ATP, carbon, CHN, and so forth. As noted in the data or remarks written in Figure 4, steps were taken though reruns, refiltration, or delineations (in cases of inadequate samples) to determine the cause of sample intercomparison errors.

The computer program will disallow the replacement or entry of a piece of data if the analysis was not scheduled, if a replicate was not noted, or if there is an analytical quality control problem. Input of any entry outside of the regular scheme is flagged in the "unmatched data card file." This flag again puts the burden on the analyst. An example of an "unmatched data card" page output is shown in Figure 5.

Each week a list of replicates is printed as shown in Figure 6. Those replicates shown with four stars require inspection by the section chief.

The printing of data that were rejected with reason by program "TECKNICON" is a recent innovation. An example of this output is shown in Figure 7. The designation, "past off," is a code indicating that the previous sample was off scale and that the present sample may be adversely affected because of washout characteristics.

Another system used in the past is a milliequivalent comparison. This comparative approach requires the complete analysis of soluble major ionic components in the water system. An output of this type is shown in Figure 8. When the cations and anions do not match, the milliequivalent balance scheme points out problems in analysis of soluble components or the possibility of the presence of an unanalyzed ionic component. The section chief must determine the action necessary to resolve problems noted in milliequivalent balance output.

Figures 4 through 8 are a few examples of quality assurance aspects incorporated through ADP into laboratory operations phases.

After 8 years experience with ADP, we have come to the following conclusions:

- ADP can be a tremendous asset in laboratory operation, especially when handling a large workload.
- Not only has ADP proved effective, its use in laboratory situations will increase.
- As in all laboratory functions, quality assurance within all phases of ADP is a first consideration.

In summary, let me turn to Rudyard Kipling's poem, "Arithmetic on the Frontier," and quote an excerpt from the poem:

> "A great and glorious thing it is
> To learn, for seven years or so,
> The Lord knows what of that and this,
> Ere reckoned fit to face the foe - "[4]

Eight years of experience with ADP has convinced us that ADP is a great tool. With ADP in hand, the laboratory manager can wage the fight against the foe (bad data).

REFERENCES

1 Chandor, A., Graham, J., and Williamson, R., *A Dictionary of Computers.* Baltimore: Penguin Book Inc., 1970, p. 27.

2 Byram, K. V., Roberts, F. A., and Wilson, L. A., "A Data Reduction System for an Automatic Colorimeter," *Proceedings No. 1, ORD ADP Workshop*, 1974.

3 Krawcyzk, D. F. and Byram, K. V., "Management System for an Analytical Chemical Laboratory," *Amer. Lab.*, vol. 5, no. 1 (1973), pp. 55-62.

4 Kipling, Rudyard. *Rudyard Kipling Verse, Definitive Edition*, Garden City: Doubleday and Company, Inc., 1973.

5 Krawczyk, D. F., Taylor, P. L., and Kee, Jr., W. D., Laboratory Sample Control and Computer Verification of Results at the Lake Huron Program Office. Paper presented at the Ninth Conference of Great Lakes Research, Illinois Institute of Technology, Research Institute, Chicago. March 29, 1966.

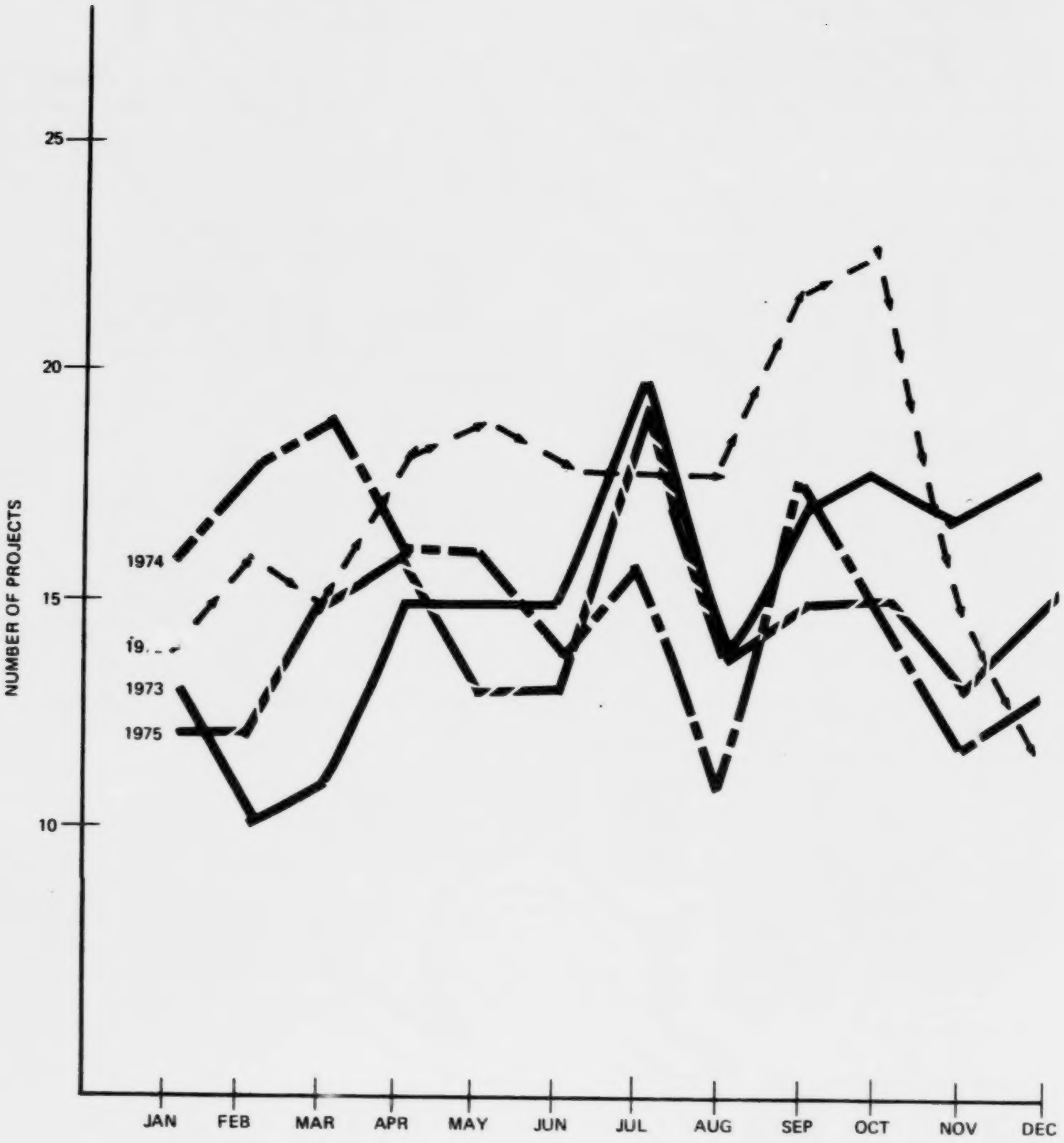


Figure 1
Projects Submitting Samples

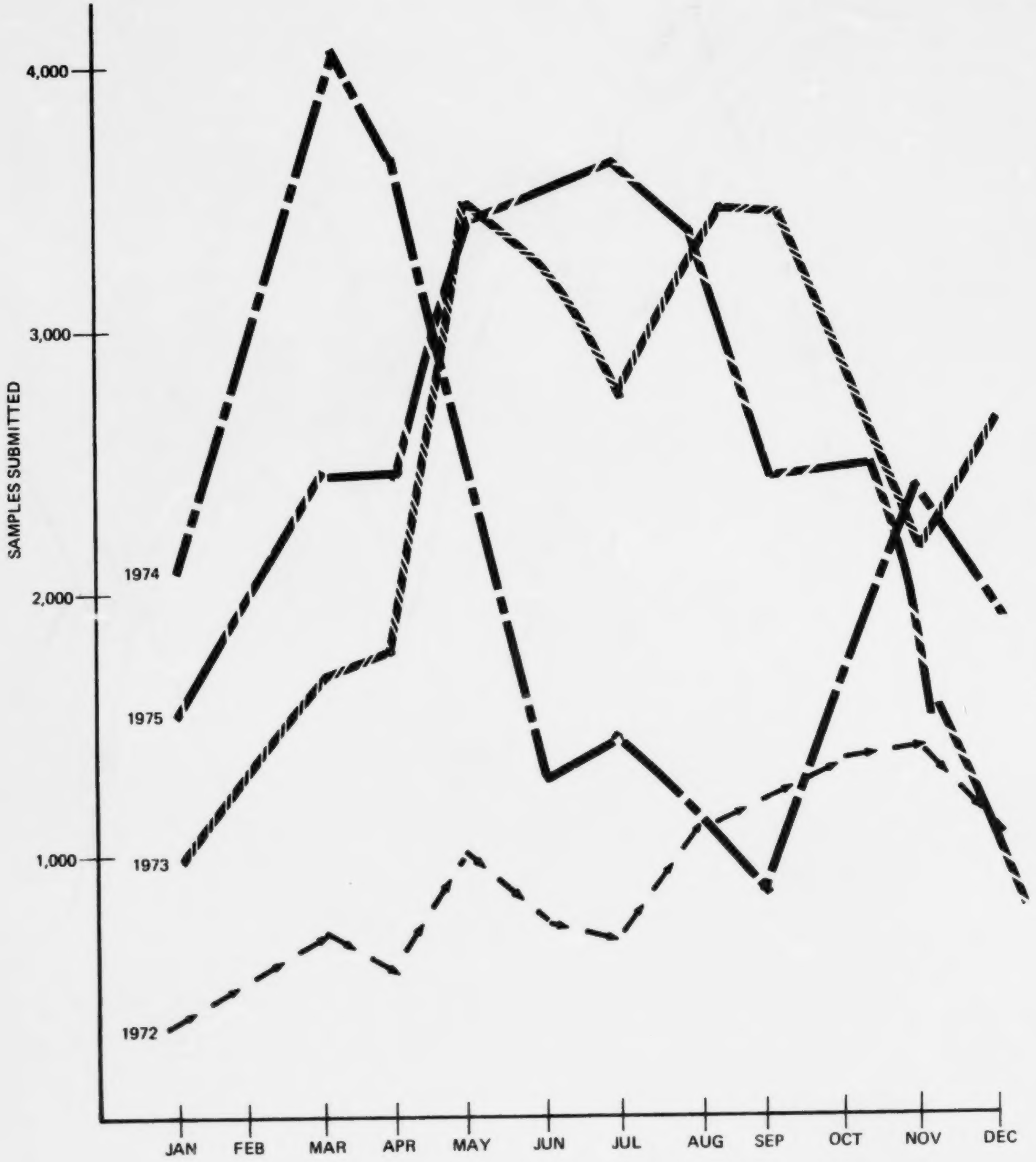


Figure 2
Samples Submitted for Chemical Analyses

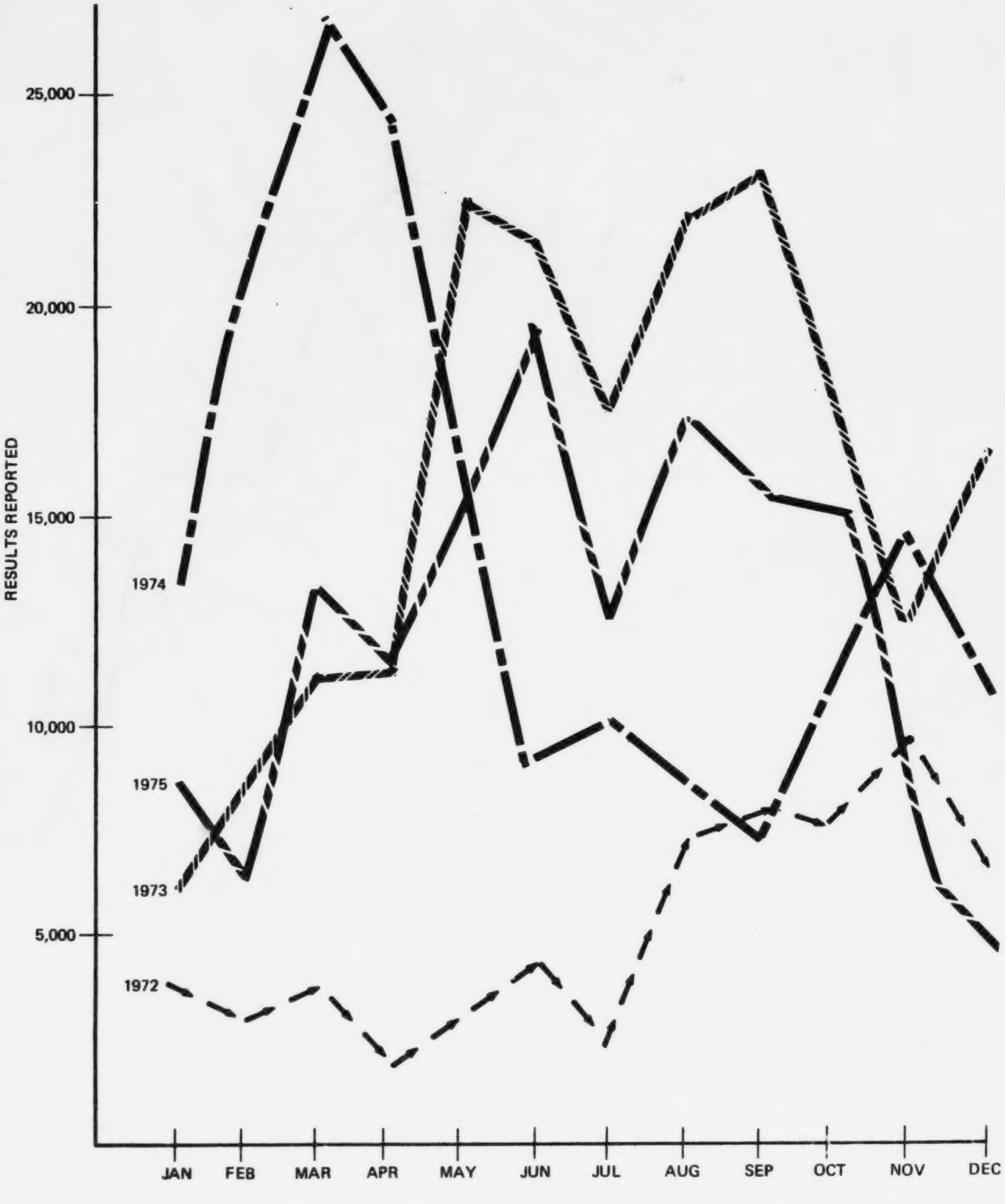
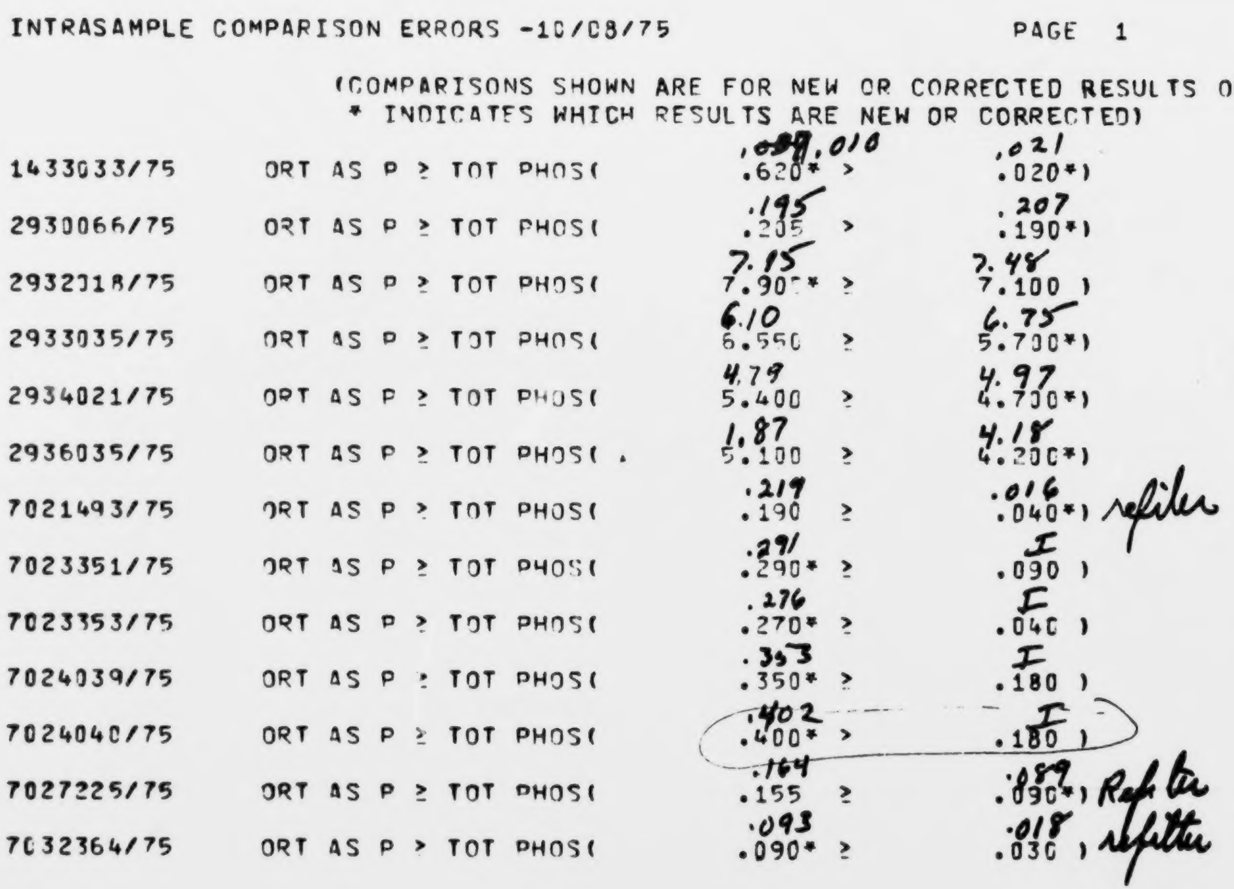
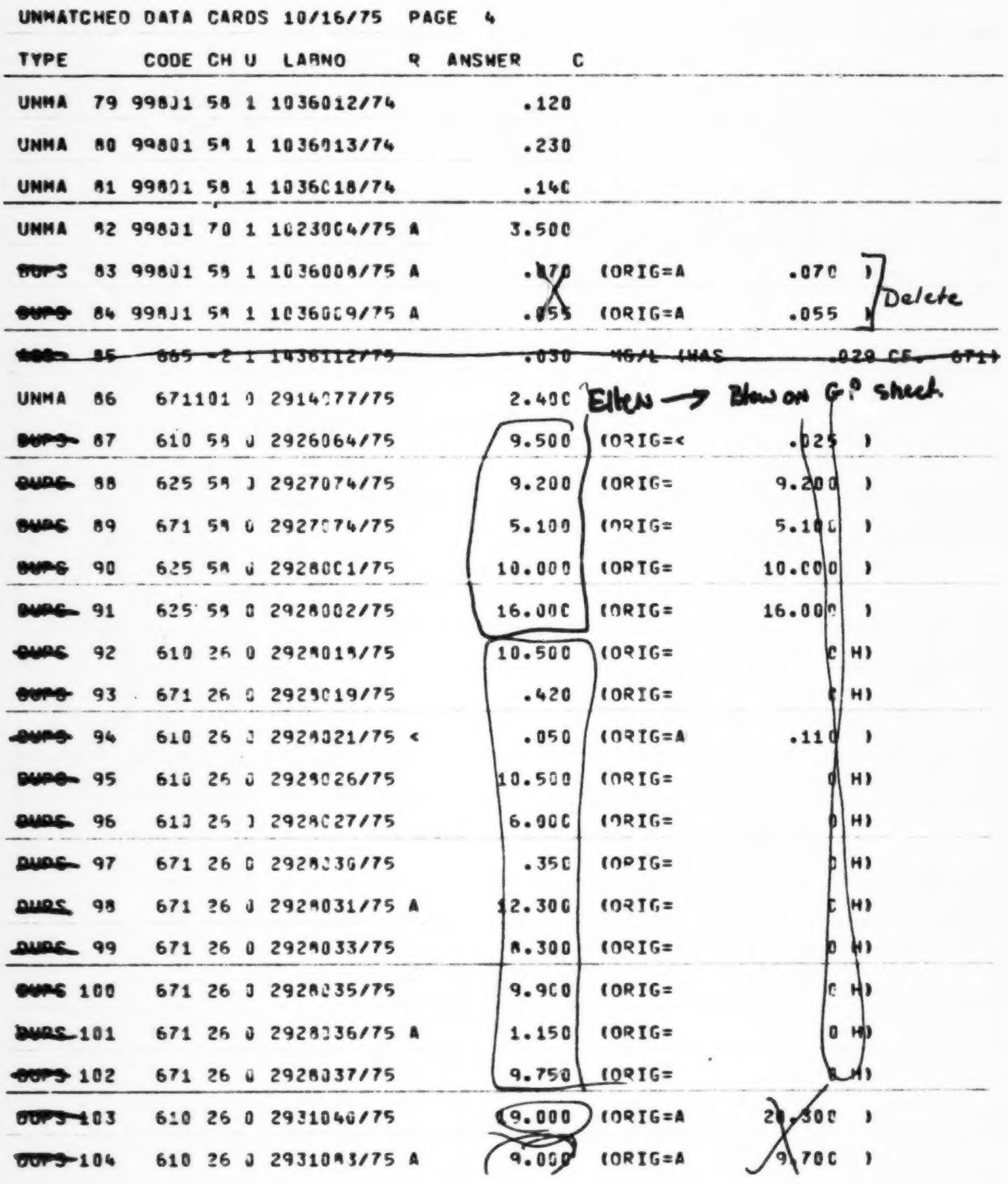


Figure 3
Results Reported to Project Leaders



```
INTRASAMPLE COMPARISON ERRORS -10/08/75                         PAGE  1

                   (COMPARISONS SHOWN ARE FOR NEW OR CORRECTED RESULTS ONLY
                    * INDICATES WHICH RESULTS ARE NEW OR CORRECTED)
                                                 ~~.039~~.010         .021
1433033/75    ORT AS P ≥ TOT PHOS(           .620* >            .020*)
                                                 .195               .207
2930066/75    ORT AS P ≥ TOT PHOS(           .205  >            .190*)
                                                 7.15               7.48
2932018/75    ORT AS P ≥ TOT PHOS(          7.900* ≥           7.100 )
                                                 6.10               6.75
2933035/75    ORT AS P ≥ TOT PHOS(          6.550  ≥           5.700*)
                                                 4.79               4.97
2934021/75    ORT AS P ≥ TOT PHOS(          5.400  ≥           4.700*)
                                                 1.87               4.18
2936035/75    ORT AS P ≥ TOT PHOS(          5.100  ≥           4.200*)
                                                 .219               .016
7021493/75    ORT AS P ≥ TOT PHOS(           .190  ≥            .040*) refilter
                                                 .291               I
7023351/75    ORT AS P ≥ TOT PHOS(           .290* ≥            .090 )
                                                 .276               I
7023353/75    ORT AS P ≥ TOT PHOS(           .270* ≥            .040 )
                                                 .353               I
7024039/75    ORT AS P ≥ TOT PHOS(           .350* ≥            .180 )
                                                 .402               I
7024040/75    ORT AS P ≥ TOT PHOS(           .400* >            .180 )
                                                 .164               .089
7027225/75    ORT AS P ≥ TOT PHOS(           .155  ≥            .090*) Refilter
                                                 .093               .018
7032364/75    ORT AS P > TOT PHOS(           .090* ≥            .030 ) refilter
```

Figure 4
Intrasample Comparison Errors Reasonable Chemical Comparisons



```
UNMATCHED DATA CARDS 10/16/75  PAGE  4
TYPE        CODE CH U  LABNO       R  ANSWER     C
UNMA  79 99801 58 1 1036012/74             .120
UNMA  80 99801 58 1 1036013/74             .230
UNMA  81 99801 58 1 1036018/74             .140
UNMA  82 99801 70 1 1023004/75 A          3.500
DUPS  83 99801 58 1 1036008/75 A           .070  (ORIG=A          .070  )
DUPS  84 99801 58 1 1036009/75 A           .055  (ORIG=A          .055  )
[illegible] 85   665  2 1 1436112/75             .030  MG/L (WAS         .029 CF. 671)
UNMA  86    671101 0 2914077/75           2.400
DUPS  87    610 58 0 2926064/75           9.500  (ORIG=<          .025  )
DUPS  88    625 58 0 2927074/75           9.200  (ORIG=          9.200  )
DUPS  89    671 58 0 2927074/75           5.100  (ORIG=          5.100  )
DUPS  90    625 58 0 2928001/75          10.000  (ORIG=         10.000  )
DUPS  91    625 58 0 2928002/75          16.000  (ORIG=         16.000  )
DUPS  92    610 26 0 2928018/75          10.500  (ORIG=              0 H)
DUPS  93    671 26 0 2928019/75            .420  (ORIG=              0 H)
DUPS  94    610 26 0 2928021/75 <          .050  (ORIG=A          .110  )
DUPS  95    610 26 0 2928026/75          10.500  (ORIG=              0 H)
DUPS  96    610 26 0 2928027/75           6.000  (ORIG=              0 H)
DUPS  97    671 26 0 2928030/75            .350  (ORIG=              0 H)
DUPS  98    671 26 0 2928031/75 A        12.300  (ORIG=              0 H)
DUPS  99    671 26 0 2928033/75           8.300  (ORIG=              0 H)
DUPS 100    671 26 0 2928035/75           9.900  (ORIG=              0 H)
DUPS 101    671 26 0 2928036/75 A         1.150  (ORIG=              0 H)
DUPS 102    671 26 0 2928037/75           9.750  (ORIG=              0 H)
DUPS 103    610 26 0 2931040/75           9.000  (ORIG=A         20.300  )
DUPS 104    610 26 0 2931083/75 A         9.000  (ORIG=A          9.700  )
```

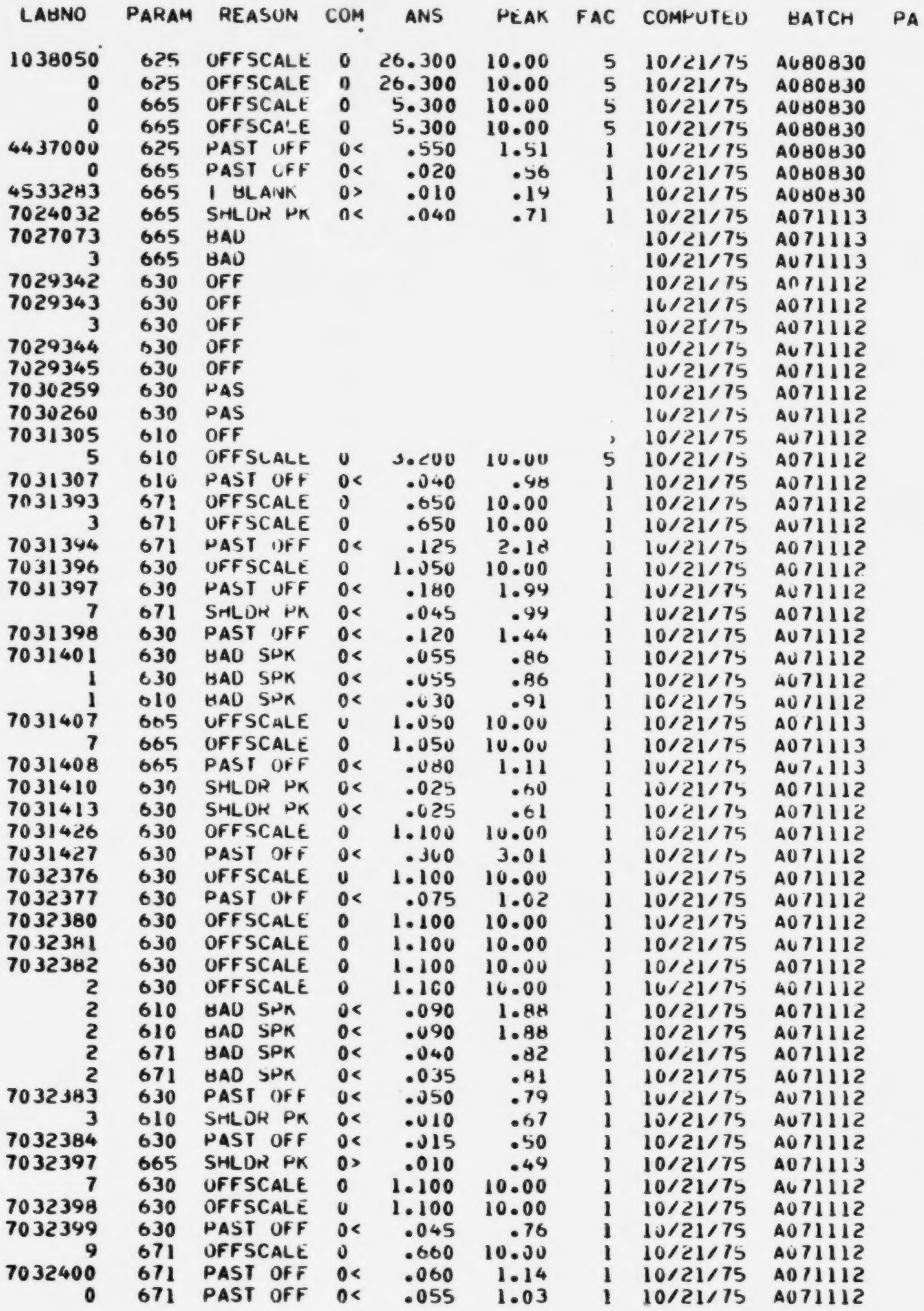
Delete

Ellen → Blow on G.P. sheet.

Figure 5
Computer Printout of Data Requiring Decisions

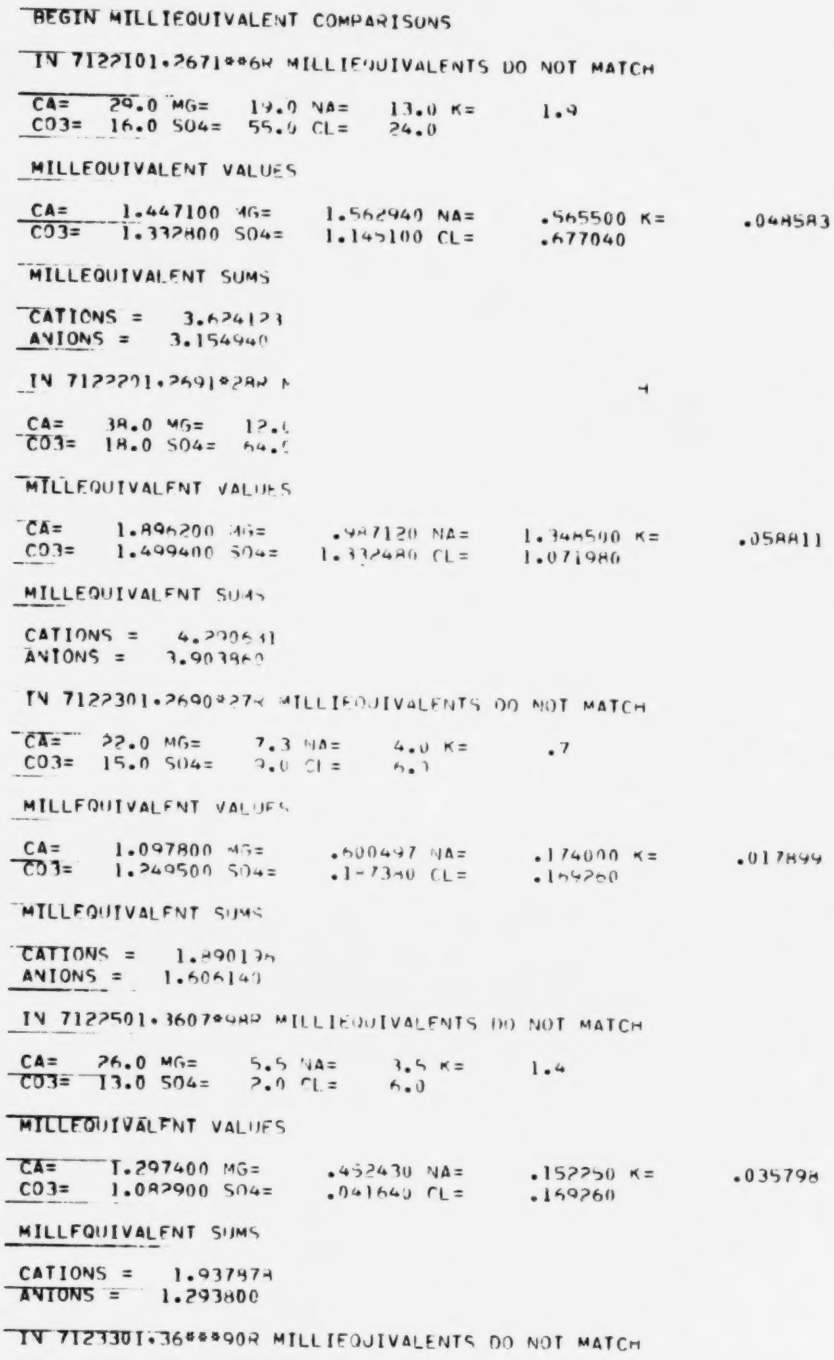
```
 REPLICATES FOR  10/22/75

         610     535001          .020         .015
                                 .0177        .0170

         610     539008          .010         .010
                                 .0091        .0091

  ****   625     539008         1.000        1.100
                                1.0025       1.0969

         630     539008S         .005S        .005
                               -0.0023      -0.0023

         665     539008          .070         .070
                                 .0667        .0700

         671     539108S         .005S        .005
                                 .0034        .0034

         625    1038002          .400         .350
                                 .3754        .3477

         665    1038002          .040         .030
                                 .0385        .0296

         671    1038102          .010         .010
                                 .0117        .0117

         610    2928034        10.500       10.500
                               10.4332      10.5924

  ****   610    2928036          .081         .094
                                 .0814        .0937

  ****   671    2928036         1.200        1.250
                                1.1912       1.2527

  ****   630    2930071        24.000       22.000
                               24.0745      22.4947

         630    2935048          .160         .160
                                 .1645        .1591

  ****   610    2936003S         .025S        .025
                               -0.0037      -0.0106

  ****   630    2936034          .065         .100
                                 .0647        .0988

         671    2937000          .120         .110
                                 .1172        .1137

         671    2937059         1.460        1.400
                                1.4171       1.4236

         630    2937060        12.600       13.000
                               12.6526      12.9701

         671    2937060         6.900        7.100
                                5.8775       7.0637

         610    2938002         1.450        1.450
                                1.4259       1.4747

         671    2938002         2.880        2.880
                                2.8662       2.8534

         630    2938012          .190         .190
                                 .1912        .1915

         630    2938017         9.200        9.300
                                9.2339       9.2953

         671    2938017         4.200        4.200
                                4.1606       4.2134

         630    2938020         4.400        4.400
                                4.4020       4.3848

         671    2938021         9.200        9.400
                                9.1591       9.3625

         610    2938022          .400         .410
```

Figure 6
Replicates With Coded Designation
(Coded-Starred Require Judgment on Acceptance)



```
  LABNO   PARAM  REASON    COM   ANS      PEAK   FAC  COMPUTED   BATCH    PA

1038050    625   OFFSCALE  0   26.300   10.00   5   10/21/75   A080830
      0    625   OFFSCALE  0   26.300   10.00   5   10/21/75   A080830
      0    665   OFFSCALE  0    5.300   10.00   5   10/21/75   A080830
      0    665   OFFSCALE  0    5.300   10.00   5   10/21/75   A080830
4437000    625   PAST OFF  0<    .550    1.51   1   10/21/75   A080830
      0    665   PAST OFF  0<    .020     .56   1   10/21/75   A080830
4533283    665   I BLANK   0>    .010     .19   1   10/21/75   A080830
7024032    665   SHLDR PK  0<    .040     .71   1   10/21/75   A071113
7027073    665   BAD                                10/21/75   A071113
      3    665   BAD                                10/21/75   A071113
7029342    630   OFF                                10/21/75   A071112
7029343    630   OFF                                10/21/75   A071112
      3    630   OFF                                10/21/75   A071112
7029344    630   OFF                                10/21/75   A071112
7029345    630   OFF                                10/21/75   A071112
7030259    630   PAS                                10/21/75   A071112
7030260    630   PAS                                10/21/75   A071112
7031305    610   OFF                                10/21/75   A071112
      5    610   OFFSCALE  0    3.200   10.00   5   10/21/75   A071112
7031307    610   PAST OFF  0<    .040     .98   1   10/21/75   A071112
7031393    671   OFFSCALE  0     .650   10.00   1   10/21/75   A071112
      3    671   OFFSCALE  0     .650   10.00   1   10/21/75   A071112
7031394    671   PAST OFF  0<    .125    2.18   1   10/21/75   A071112
7031396    630   OFFSCALE  0    1.050   10.00   1   10/21/75   A071112
7031397    630   PAST OFF  0<    .180    1.99   1   10/21/75   A071112
      7    671   SHLDR PK  0<    .045     .99   1   10/21/75   A071112
7031398    630   PAST OFF  0<    .120    1.44   1   10/21/75   A071112
7031401    630   BAD SPK   0<    .055     .86   1   10/21/75   A071112
      1    630   BAD SPK   0<    .055     .86   1   10/21/75   A071112
      1    610   BAD SPK   0<    .030     .91   1   10/21/75   A071112
7031407    665   OFFSCALE  0    1.050   10.00   1   10/21/75   A071113
      7    665   OFFSCALE  0    1.050   10.00   1   10/21/75   A071113
7031408    665   PAST OFF  0<    .080    1.11   1   10/21/75   A071113
7031410    630   SHLDR PK  0<    .025     .60   1   10/21/75   A071112
7031413    630   SHLDR PK  0<    .025     .61   1   10/21/75   A071112
7031426    630   OFFSCALE  0    1.100   10.00   1   10/21/75   A071112
7031427    630   PAST OFF  0<    .300    3.01   1   10/21/75   A071112
7032376    630   OFFSCALE  0    1.100   10.00   1   10/21/75   A071112
7032377    630   PAST OFF  0<    .075    1.02   1   10/21/75   A071112
7032380    630   OFFSCALE  0    1.100   10.00   1   10/21/75   A071112
7032381    630   OFFSCALE  0    1.100   10.00   1   10/21/75   A071112
7032382    630   OFFSCALE  0    1.100   10.00   1   10/21/75   A071112
      2    630   OFFSCALE  0    1.100   10.00   1   10/21/75   A071112
      2    610   BAD SPK   0<    .090    1.88   1   10/21/75   A071112
      2    610   BAD SPK   0<    .090    1.88   1   10/21/75   A071112
      2    671   BAD SPK   0<    .040     .82   1   10/21/75   A071112
      2    671   BAD SPK   0<    .035     .81   1   10/21/75   A071112
7032383    630   PAST OFF  0<    .050     .79   1   10/21/75   A071112
      3    610   SHLDR PK  0<    .010     .67   1   10/21/75   A071112
7032384    630   PAST OFF  0<    .015     .50   1   10/21/75   A071112
7032397    665   SHLDR PK  0>    .010     .49   1   10/21/75   A071113
      7    630   OFFSCALE  0    1.100   10.00   1   10/21/75   A071112
7032398    630   OFFSCALE  0    1.100   10.00   1   10/21/75   A071112
7032399    630   PAST OFF  0<    .045     .76   1   10/21/75   A071112
      9    671   OFFSCALE  0     .660   10.00   1   10/21/75   A071112
7032400    671   PAST OFF  0<    .060    1.14   1   10/21/75   A071112
      0    671   PAST OFF  0<    .055    1.03   1   10/21/75   A071112
```

Figure 7
Computer Output of Rejected Data



```
BEGIN MILLIEQUIVALENT COMPARISONS

IN 7122101.2671**6R MILLIEQUIVALENTS DO NOT MATCH

CA=   29.0 MG=   19.0 NA=   13.0 K=     1.9
CO3=  16.0 SO4=  55.0 CL=   24.0

MILLEQUIVALENT VALUES

CA=     1.447100 MG=    1.562940 NA=     .565500 K=      .048583
CO3=    1.332800 SO4=   1.145100 CL=     .677040

MILLEQUIVALENT SUMS

CATIONS =   3.624123
ANIONS =   3.154940

IN 7122201.2691*28R M                                       H

CA=   38.0 MG=   12.
CO3=  18.0 SO4=  64.

MILLEQUIVALENT VALUES

CA=    1.896200 MG=     .987120 NA=    1.348500 K=      .058811
CO3=   1.499400 SO4=   1.332480 CL=    1.071980

MILLEQUIVALENT SUMS

CATIONS =   4.290631
ANIONS =   3.903860

IN 7122301.2690*27R MILLIEQUIVALENTS DO NOT MATCH

CA=   22.0 MG=    7.3 NA=    4.0 K=      .7
CO3=  15.0 SO4=   9.0 CL=    6.0

MILLFQUIVALFNT VALUES

CA=    1.097800 MG=     .600497 NA=     .174000 K=      .017899
CO3=   1.249500 SO4=    .187380 CL=     .169260

MILLFQUIVALFNT SUMS

CATIONS =   1.890196
ANIONS =   1.606140

IN 7122501.3607*98R MILLIEQUIVALENTS DO NOT MATCH

CA=   26.0 MG=    5.5 NA=    3.5 K=     1.4
CO3=  13.0 SO4=   2.0 CL=    6.0

MILLEQUIVALFNT VALUES

CA=    1.297400 MG=     .452430 NA=     .152250 K=      .035798
CO3=   1.082900 SO4=    .041640 CL=     .169260

MILLFQUIVALFNT SUMS

CATIONS =   1.937878
ANIONS =   1.293800

IN 7123301.36***90R MILLIEQUIVALENTS DO NOT MATCH
```

Figure 8
Special Computer Program to Compare Chemical Balance in Water Sample

THE CLEANS/CLEVER AUTOMATED CLINICAL LABORATORY PROJECT AND DATA MANAGEMENT ISSUES

By Sam D. Bryan

INTRODUCTION

As participants of this second ORD ADP workshop, we were asked to orient our discussions around ADP issues. An issue involves a decision of importance; therefore, this paper will present a brief discussion of some data management decisions important to the CLEANS/CLEVER project. Some of these decisions have already been made, but others will not be made for the next several months. Hopefully, this discussion will be useful to others who are now, or will be, faced with making similar decisions. The opinions stated are the author's and are not necessarily shared by anyone else.

CLEANS/CLEVER SYSTEMS OVERVIEW

The Clinical Laboratory Evaluation and Assessment of Noxious Substances (CLEANS) program will be located in the present EPA Clinical Studies facility at the University of North Carolina in Chapel Hill. The Clinical Laboratory Evaluation and Validation of Epidemiologic Research (CLEVER) program will be located in sophisticated mobile laboratories that will travel from a home base in Chapel Hill to various locations across the United States.

The CLEANS program will be conducted using two large exposure chambers and an adjoining area for a computerized physiological data acquisition system and a pollutant control system. Using these self-contained exposure laboratories, CLEANS scientists can expose individuals for extended periods of time to air pollution conditions that are the same as those found in urban areas.

Until the CLEANS program, clinical studies have been performed only during brief 2 to 6-hour exposures to pollutants. Such short-term exposures provided only limited information on the interaction between pollutants and human physiological systems. Longer term exposures, particularly when atmospheric conditions can be precisely simulated, will greatly increase knowledge about health effects of airborne pollution.

The data system will enable the staff of the Clinical Studies Division to make more frequent and more precise measurements of human physiological reactions and to determine physiological function decrement and recovery over extended periods of time.

The controlled clinical laboratory study of air pollutants and other environmental stress on humans will provide health effects data for:

- Continued evaluation of existing air quality standards and potentially noxious substances
- Establishment of short-term standards for new air pollutants
- Establishment of standards for odors, noise, and microwave radiation.

Significantly, these programs will provide information necessary to evaluate the adequacy of current ambient air quality standards.

The two clinical environmental laboratories will have beds and bathrooms as well as televisions and telephones. Food, clean clothing, and linens will be supplied through a catering service to make the laboratories fully inhabitable both for short and extended stays. The environment of each exposure laboratory will be controlled (either manually, or automatically by the computer) for temperature, humidity, lighting level, and pollutant gas and aerosol concentration. The process control computer also will be used to record environmental parameters and operational status of the measurement instruments. The laboratories also will be equipped with an array of cardiopulmonary exercise instrumentation and an associated physiological data acquisition system. Each exposure laboratory will contain complete examining and testing equipment for the human research subjects.

For the CLEVER program, two mobile van systems will be used to study cardiopulmonary functions of humans exposed alternately to high and low ambient air pollution levels in their home environment in the United States. A standard 34-foot self-contained motor home will be the shell of the CLEVER mobile laboratory. By eliminating the standard cooking, sleeping, and living space areas in the mobile unit, room will be provided for the large array of equipment needed. Behind the driver's area, the mobile unit will contain areas for reception and interview, test preparation, subject examination, housing of the computer, medical instrumentation, and exercise equipment.

The CLEVER program's mobile laboratory will have physiological measurement equipment identical to that housed in the stationary CLEANS laboratory in Chapel Hill. Each system is designed, constructed, and instrumented to ensure that data obtained will be directly comparable between the two facilities.

The mobile facility will be used to obtain and evaluate clinical data from epidemiological studies in various population study areas.

The CLEVER mobile laboratory will travel to the home areas of populations being studied to gather data on environmental pollutants effects on human health. The primary mission of the mobile laboratory will be pulmonary function and cardiovascular performance measurements. In addition to this capability, general medical histories will be taken, physical examinations made, and biological specimens stored. Future plans call for expansion of the CLEVER mobile laboratory capabilities to include other noninvasive physiologic measurements.

Computer Sciences Corporation (CSC) has been contracted to design and build the fixed facility as well as the two mobile systems. EPA expects to begin using these systems in 1976. Some of the work discussed below is original work by CSC.

COMPARISON TO OTHER EPA LABORATORY AUTOMATION PROJECTS

Many of the issues of this project will be found in other laboratory automation projects. The paradigm is a sensing instrument connected to an analog-to-digital converter, connected to a computer, which displays, analyzes, and stores the sensed data. It is the same in this project as it is, for example, in the chemical laboratory case. This project, perhaps, differs quantitatively from the chemical laboratory case by interfacing a relatively large number of different instruments (about 12 presently) and by sampling the analog outputs of some instruments at a relatively high rate (e.g., 500 Hz for the ECG amplifiers). This clinical system also is highly integrated so that the system operator can interact with a complex array of instruments and perform elaborate subject testing protocols through a single display screen and a single customized keyboard.

In general, this clinical system is very large, complex, and nonstandard. The online acquisition and control software alone required about 30 man-years of programing. This effort, understandably, had to be achieved through a contractor. The in-house versus contract issue is extremely important, but will not be discussed here.

The complexity of the system stems from the need to do a number of tasks concurrently. For example, it must sample, display, and store ECG signals, do simple analyses on them, and control the speed and elevation of an exercise treadmill.

There are both nonstandard hardware and nonstandard software features. Special hardware interfaces are required between the medical instruments and the computer since there is generally a disparity in the output voltage ranges of the medical instruments and the A/D input range of the computer. Special hardware was also designed to interface the human operator with the rest of the system. This special equipment takes the form of a customized keyboard and display. It was considered very important to make the operator's interaction with the system as streamlined as possible in order to process a high throughput of subjects. Interaction was made as simple and foolproof as possible in order to simplify operator training and to reduce operator mistakes. The issue of using off-the-shelf versus custom-built hardware devices and interfaces is another important issue; however, it is outside the scope of this paper.

The system uses nonstandard operating systems as well. In particular, the PDP-11, RSX-11A, and DOS operating systems were modified to coexist on the RK05 disk, to share disk files, and to pass control of execution back and forth. The gas monitoring and control computer uses the RT-11 operating system.

Another distinctive feature of this laboratory automation project is the overwhelming emphasis placed on subject safety. This is apparent mainly in the pollutant control system, where elaborate concentration-range checks and performance-status checks are made.

High reliability of the total system is also of paramount importance, since downtime will be so expensive both in terms of its possible damaging effect on experimental protocols and on the time of test subjects and technicians that it would waste.

We are faced with pushing the state-of-the-art in some cases, where an attempt is made to automate a previously manual task. Such ventures incur time and money risks. Automating versus not automating is another important issue that is outside the scope of this paper.

Although some of the issues we face in the CLEANS/CLEVER project differ at least quantitatively from issues faced in other laboratory projects, most of the data management issues discussed below are relevant.

THE ISSUE OF ONLINE VERSUS OFFLINE REPORTS

In the physiological system, which is considered here, many of the tests require subject cooperation. For example, in the forced vital capacity spirometry test, the operator exhorts the subject to exhale as fast and as completely as possible. This maximal effort is used as a norm to make comparisons to other subjects, or comparisons to the same subject at different times. Since a subject "maneuver" is often inadequate, usually multiple maneuvers are required. This is one important reason for seeing the results of the tests online, so that the operator can direct the subject to try again if poor performance is indicated. To store spirometer output voltages blindly on an analog tape, for example, would either require the subject to perform a large number of maneuvers, which is not desirable, or to perform just a few maneuvers in the uncertain hope that he has done his best, which is an even less desirable alternative.

The subject is just one of several components of the system with which something can go wrong. In the case of spirometry, for example, any of the following can be faulty: the spirometer bellows, transducer, amplifier, signal conditioning circuitry, A/D converter, computer hardware, or software. The ability to display the spirometer signal and the numeric values derived from it online gives the operator the necessary opportunity to note a problem on the spot and thereby prevent erroneous data from being recorded. Although this is an important capability in the chamber facility, it is even more important in the vans, where it is not always feasible to have a subject come back for retesting if something went wrong.

The other physiological measurements also require on-the-spot abbreviated reports on overall data acquisition status for the operator. Offline reports are appropriate once successful acquisition of the data has been assured. The longer, more fine-grained reports on physiological response required for subsequent correlation with the pollutant data should be done off-line. This is interrelated with the following issue.

THE ISSUE OF COMPUTING ON THE LABORATORY MINICOMPUTER VERSUS THE CENTRAL MAXICOMPUTER

This question involves which programing tasks should be performed on the laboratory minicomputer and which ones on the central maxicomputer. The answer, as usual depends on the task.

Some tasks must be done on the laboratory minicomputers. These are the real-time tasks involving fast sampling rates or special laboratory devices that cannot be supported at a terminal, or tasks requiring such a large percentage of uptime, as in the pollutant control system, that a dedicated minicomputer is required both technically, because simpler machines are more reliable and managerially, because dedicated administrative control is required for a critically important function.

Some tasks must be done on the central maxicomputer. These are tasks that involve very large programs, such as the sophisticated statistical packages, or that involve very large online files, such as those required for efficient sorting or online queries of a large data base.

There are also the tasks that theoretically can be programed on either the minicomputer or the maxicomputer. The turnaround time requirement can swing the decision to the use of one or the other. If turnaround time on the order of 48 hours is acceptable, then the maxicomputer is appropriate. The minicomputer should be left as free as possible for tasks that it is uniquely suited to run. To arbitrarily add tasks to the small system eventually would create, unnecessarily, a small computer center with the attendant problems of scheduling, tape and disk library management, more operators, supplies, and maintenance problems. It is possible that evolution of this type of operation is unavoidable, but a special effort should be made to avoid arbitrary loading of the minicomputer system. If a task can be performed on the maxicomputer (and if there is no fast turnaround requirement), a number of advantages result: 1) the large machine is generally more accessible both for program development and for production runs since it timeshares a large number of terminals; 2) there is generally more manpower available for programing assistance on the maxicomputer than on the minicomputer; 3) program development is generally easier on the maxicomputer due to more core availability and a wider variety of high-level languages; and 4) there exists greater availability of large general purpose statistical and data management packages.

There is the final case to consider; that of a task which could be programed on either machine, but which has a fast turnaround requirement (e.g., 4 hours). Such tasks involve status reports based on the offline processing of the "history" tapes output by the physiological and pollutant system. Examples are: 1) trend-plotting of recent physiological measurements that would be used both for subject safety supervision and for quality assurance of the entire system, and 2) limit-checking of both physiological and pollutant variables that would be used for the same purposes. These are examples of status checks used to protect the subject's safety and to ensure the acquisition of valid data. Such status information may indicate a need for fast remedial action, thus, fast turnaround is mandatory.

The conservative approach would be to perform these tasks on the labortory machine where fast turnaround is more likely to be achieved on a 24-hour, 7-day-a-week basis due to the factors, mentioned above, of small machine reliability and local administrative control.

Even if the maxicomputer were 100 percent reliable and accessible, it still would not necessarily be the proper choice for these tasks. The only way of achieving fast job turnaround using the maxicomputer would be to enter the jobs through an RJE terminal at the remote laboratory site and to get the listings back over the RJE terminal. This is because our closest laboratory computer, which is in the fixed exposure facility, is over 10 miles away from the central maxicomputer site, and shipping tapes and listings back and forth would be too time-consuming. This introduces two difficulties. First, too often the communications link either is down or drops connections in midtransmission. Second, at least for the foreseeable future, one of the data acquisition minicomputers in the fixed exposure facility will be used as the RJE terminal. If we have RJE capability in the mobile vans, it will certainly be through the onboard minicomputer. One important goal is to free the minicomputer by using the maxicomputer. However, the simple report-generating programs that we are considering here, trend-plotting, and limit-checking, are I/O-bound, and the minicomputer would be tied up sending data over a relatively slow communications line (e.g., 2000 baud), waiting some time for the job to be run, and then waiting for the report to be sent back for printing at a rate much slower than the minicomputer's line printer rate. So the minicomputer is tied up longer for this kind of task in trying to use the maxicomputer than it would be in simply running the task on the minicomputer (provided enough online file storage is available). The possiblity of running an RJE task concurrently with the physiological applications tasks is not practicable since the operating system under which the applications tasks run cannot support continuous multitasking.

The most conservative, and currently the most favored, approach is to develop these trend-plotting and limit-checking programs on the maxicomputer as well. There are several advantages to this: 1) a trend program is already required on the maxicomputer to do plots of long trends which cannot be done on the laboratory minicomputer due to limited online storage; 2) the feasibility of using the maxicomputer and the RJE in a fast-turnaround report-generating mode could be tested this way and, of course, this is the only sure way to know if that approach can work, and is superior to armchair speculation; and 3) perhaps most importantly, redundant maxicomputer analysis could be used by the quality assurance supervisor (see next section) as a cross-check against the values reported on the minicomputer. To elaborate on 3 above, this project has so many components that might fail and its resultant data is so important, that redundancy can be vital. Just recently, in a large study in the Clinical Studies Division, a problem was brought to light by redundant analysis, saving the investigators from reporting possibly incorrect results.

These status-checking programs were not included in the current CLEANS/CLEVER contract scope of work. The present plan, subject to various approvals, is to hire programers under an operations and maintenance contract to perform this programing. It is possible that in the not-so-distant future, additional online file storage may be required; and in the longer range, another CPU may be required depending on the workload imposed on the current configuration by such requirements as expanded applications programs, RJE, and onsite status checking.

THE ISSUE OF QUALITY ASSURANCE

The most critically important data management function is quality assurance of valid data. The end product of the combined online and offline systems is reports by the clinical investigator on the relationship of physiological response to pollutant dose. The importance of the data underlying these reports is obvious.

However, the size and complexity of the system demands a special effort to ensure that everything is working properly to produce valid data. In fact, both the online acquisition system and the offline data management system are complex. The difficulties that could arise online were mentioned previously with the need for online reports, using the spirometry measurement as an example.

In the offline case there are fewer potential hardware problems, but there are some data flow logistics problems. First, there are many possible sources of data. They include:

- The physiological system history tape
- The controlled pollutant system history tape
- Handwritten physiological values (when the automatic system is down)
- Handwritten pollutant values (when the automatic system is down)
- Medical questionnaire form
- Microbiological and metabolic laboratory reports
- Handwritten pollutant values at the mobile van site
- Information from various operators' log books.

For some of these sources of data, coding, keypunching, and verification are required. Also, for each of these sources of data, there must be a listing and edit program. When data are detected to be in error, the edit program is used to correct the data where possible or to purge it where correction is impossible. Once remedial action has been taken, its effect must be verified. Subsequently, merging with earlier data and other forms of data must be performed, and so on. These steps finally lead to a cleaned up data base. Thus, the data flow through many stations and in a variety of forms, and the coordination of these activities presents a real challenge.

Presently, it is our feeling that a single individual should be responsible to follow the daily flow of data full-time to ensure that a clean data base actually results. This "quality assurance supervisor" also should have knowledge of the status of the instrumentation calibration, standardize testing procedures, and audit checks of the system. This supervisor will serve as a liaison among the clinical, operations, programing, and clerical personnel. He will have the important responsibility of responding to systems-oriented questions from any of these groups.

CONCLUSION

This paper has summarized some of the technical and administrative data management decisions that have been, or are in the process of being, made by persons involved in the CLEANS/CLEVER project. We appreciate having the opportunity in this forum to share our deliberations with fellow ORD computer users.

REQUIREMENTS FOR THE REGION V CENTRAL REGIONAL LABORATORY (CRL) DATA MANAGEMENT SYSTEM

By Billy Fairless

Laboratory scientists and supervisors must perform data management for the following procedures:

- Analyzing samples in a timely manner
- Observing recommended holding times for perishable parameters
- Maintaining a balanced workload for laboratory personnel
- Allowing time to identify and correct inaccurate data
- Performing in an overall efficient manner.

Prior to understanding the minimum data management requirements for a service laboratory such as the CRL, it is necessary to understand laboratory operation. We will follow a survey from inception to completion in this paper.

First, a survey is programed to satisfy a stated purpose, and a project officer (PO) is assigned to it. The PO specifies the numbers and kinds of samples to be collected and the parameters to be analyzed for each sample. He follows existing quality assurance guidelines for items such as reagent blanks, sample preservatives, bottle types, and sample volumes. In Region V, a computer technician establishes the basic data form for the survey (see Figure 1) using existing software and the OSI computer. The technician adds station identifying information (latitude-longitude, river mile, etc.) as necessary, and the finished data is input into STORET.

A copy of the basic data form is given to the data samplers prior to sample collection. The samplers then establish their travel plans, arrange for sampling bottles, sample preservatives, proper labels, shipping of collected samples, and purchase of ice or dry ice while in the field. They wash and label all sampling bottles as necessary, collect and preserve the requested samples, and complete all field measurements for parameters such as temperature, wind direction, precipitation, pH, and turbidity. Note that a sample usually will consist of at least seven different bottles and frequently will contain over ten different bottles. Many bottles are required because the parameters are preserved differently; obviously a bottle preserved with nitric acid (for metals) could not be analyzed for nitrates. A sample containing seven bottles is shown below.

Sample: 75,-19876 -

Preservative	Parameters
No preservative	pH, cond., solids, BOD, alk, etc.
Nitric acid	Metals except Ag and W
Sulfuric acid	NO_3, phos, COD
NaOH	Cyanide
$CuSO_4/H_3PO_4$	Phenols
Ice	Organics
Formalin	Biological

The CRL uses 24 different methods to preserve samples and routinely analyzes for over 200 different parameters. In FY 1974 the average number of analyses per sample was 40.

When samples from the field arrive at CRL, they are received in the laboratory shipping and receiving area where label data are checked against information on the basic data form. The form is obtained from computer by shipping and receiving personnel and includes field data entered by samplers upon return to the field office. It is necessary to correct the information on the basic data form either because some samples are broken, others are not collected, or more samples are taken than originally planned. Samples are then divided according to the laboratory section that will perform the measurements (inorganic, metals, organic, biology) and copies of appropriate pages of the corrected basic data form are given to the section chiefs.

Up to this point in the operation, all samples are kept together as a survey group. Once in the laboratory, however, efficiency dictates that they be mixed with samples from other surveys. This mixing is one of the critical differences between the operation of a high-production service laboratory and that of a research laboratory. Mixing is required because preparation time to perform an analysis is approximately 2 hours and shutdown time is 1 hour. Therefore, 3 hours are required to obtain the first parameter concentration. After this, all remaining samples are analyzed at rates between 20 and 1,200 analyses per hour. Thus, once an analysis system is set up and running, one should analyze all samples in the laboratory requesting that parameter. If samples are missed, or if some samples are not analyzed properly, a minimum of 3 hours is required the following day to complete the requested parameter work. Since the CRL employs only 20 bench chemists and runs over 200 different parameters, each chemist is responsible for an average of 10 different parameters. The Surveillance and Analysis Division requires a 14-calendar-day turnaround time which gives each chemist only one day for each assigned parameter. Therefore, when samples are not analyzed on the proper day, makeup work must be done using personnel from another group. This results in lower quality data and wasted resources. In summary, although mixing creates a serious data management problem, it permits us to operate from 300 to 500 percent more efficiently than we could by handling all samples in their original groups.

As parameters are completed, the bench chemists report the results to the section chiefs, who perform "two parameter" quality assurance audits when possible. When all measurements assigned to the section have been completed and they appear to be correct, pages of the basic data form with handwritten results are given to the computer operator. The operator enters the results into the computer using a low-speed terminal, a key punch, or an optical card reader. When we have gained more experience with the system, we believe the chemist may be able to enter his own data using the optical card reader.

When all sections have reported their results for a given survey, the PO is notified by phone. He retrieves a copy of the basic data form from the computer, submits all data to an automatic quality assurance audit, reviews the results for reasonableness, and refers questions back to the appropriate section chief. After the PO has accepted the results, he writes his report and directs that the data on the basic data form be entered directly into STORET by the computer technician. Later, a retrieval is made during the weekly update of that system to ensure that all the data were placed in STORET.

Given the above information, the minimum requirements from a laboratory data management system can be easily summarized. The system should provide:

- A summary of work to be done by date, survey, section, and parameter for management short-term planning. Figure 2 is an example of a workload listing by parameter for the inorganic section.
- A real-time listing of in-house samples to be analyzed for a given parameter. For example, if a chemist is running mercury, the computer should identify all in-house samples requiring mercury.
- A report giving the status of each survey including: due date, analyses completed, analyses being run, and analyses not started. Figure 3 is an example of how the CRL collects and reports this data manually.
- A summary report of work completed per unit of resource expended so that long-term plans can be made. See Figure 4 for the computer output desired.

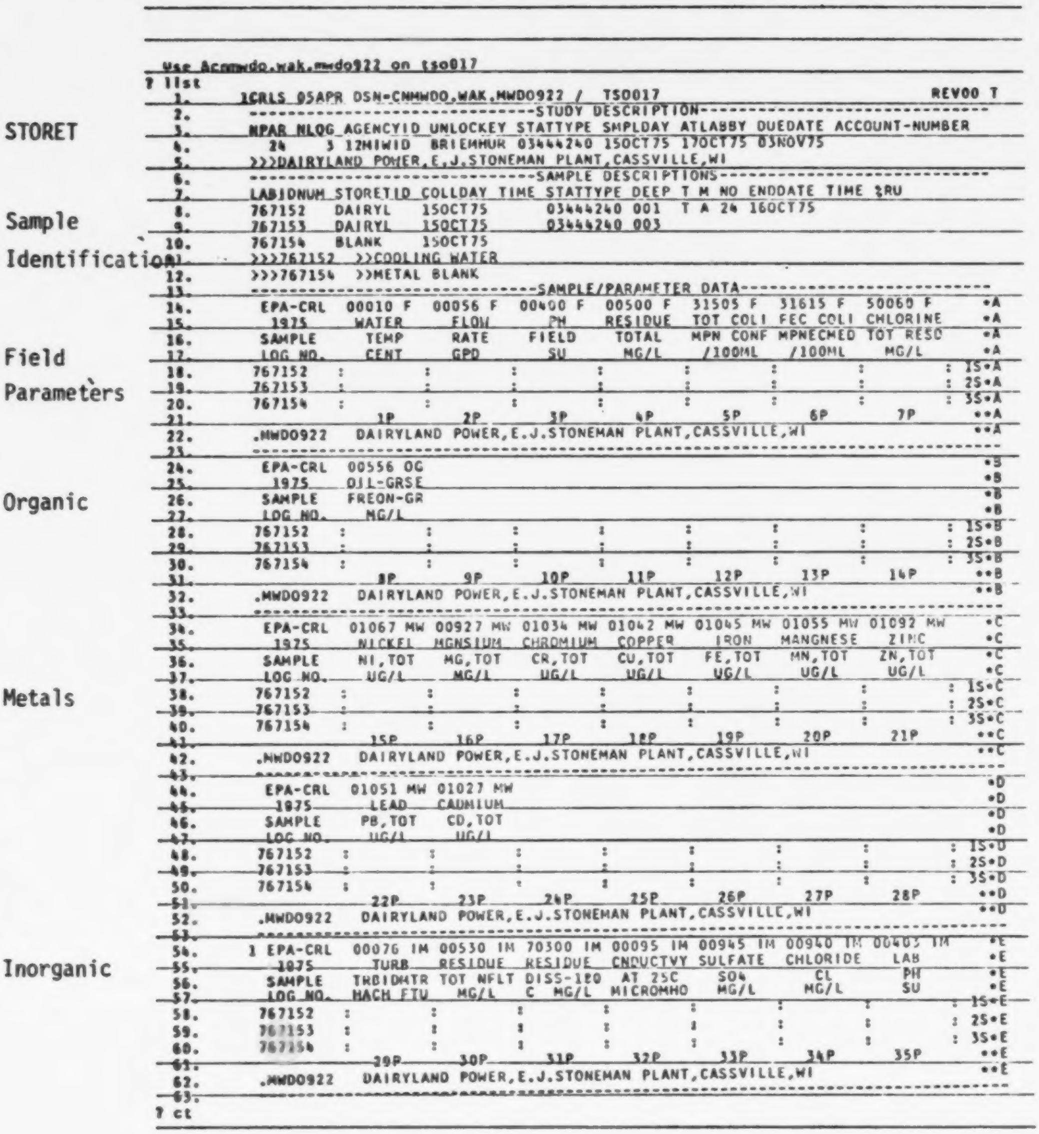
Figure 5 shows a graph of the CRL requested workload as a function of time for this fiscal year. Please note that the first spike represents a rate of work that would require a staff of 100 chemists. Since the holding times used by the CRL permit us to hold some samples for 1 week, others for 2 weeks, and still others for longer time periods, we are able to analyze all of them with our 23-man staff without losing samples (if we are not asked to do additional work the following week). When we obtain large numbers of samples on consecutive weeks, however, as shown in the last spike in Figure 5, we are forced to discard unanalyzed samples when the holding times expire. Workload spikes of this nature usually occur because work request information cannot be processed in the time necessary to effect a change in sampling plans.

In addition to knowing the total number of measurements to be made, it is essential that we know our workload on a parameter-by-parameter basis, to ensure that all perishable samples are analyzed first. Figure 2 is a backlog listing from our inorganic section for the week of October 17, 1975. Dr. Carter had just over 1,000 phosphorous, 361 TOC, 459 mercury, and many other analyses to complete. However, he also had 22 bromide analyses and, since we do not run bromide on a regular basis, we knew these 22 analyses would require 2 man-days; almost as much time as the 459 mercury analyses would require. As you can see, if we are to avoid discarding samples, it is essential that workload listings on a parameter basis be available in real time to the section chiefs so that section personnel can be used effectively. Presently, we are using at least three positions to provide this information in a timely manner.

Figure 3 is one page of a weekly publication we print summarizing the status of each survey. The left column identifies the survey by name, the submitting office, the computer data set number, and the beginning and ending sample log numbers. The next three columns record key dates and estimates of the work required by each section. The fifth column is titled "No. Analyses Requested." The last section permits us to identify quickly which parameters are completed (0) for a particular survey and which remain to be done (O). Often, when program deadlines are approached, a PO will request an incomplete data set and begin his report; he will finish it as the last parameters are completed.

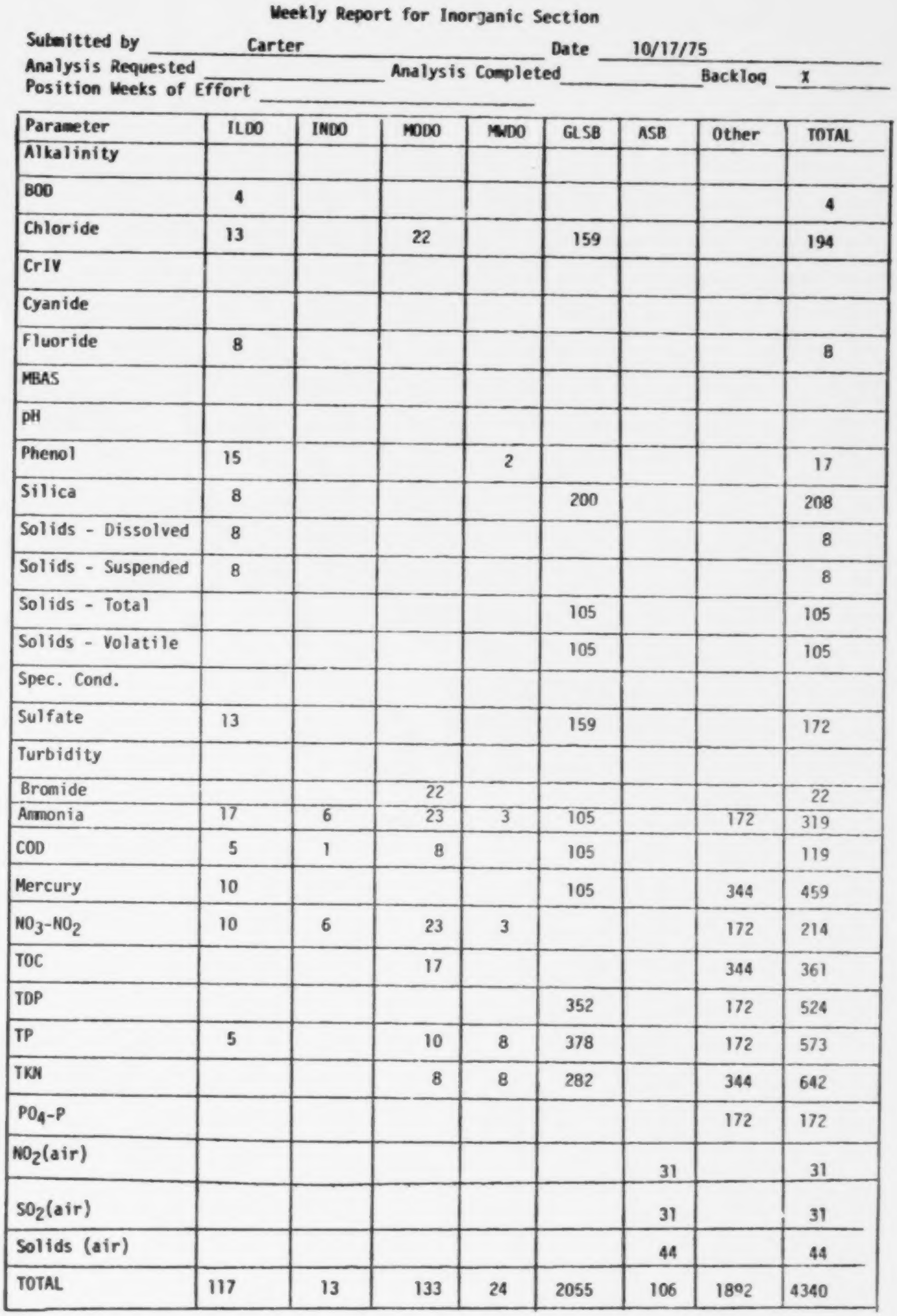
Figure 4 is a summary used at the CRL for long-term management purposes, such as estimating resources required for different work-plan options, the proper balance of personnel among the sections, and ability to participate in national or other large projects. For this and for all other outputs, we define an analysis as a concentration value reported to someone else. Therefore, these numbers probably describe between 10 to 30 percent of all data we would like to have automated into a complete Laboratory Data Management system. Figure 6 is the output from Lab-Label which tracks the progress of each survey. The example is from the Indiana District Office (INDO) file. At the end of each year this file will contain a summary of the work completed by INDO. At any time during the year it can be used to spot bottlenecks. The report is generated automatically by Lab-Label and requires very little computer operator time.

As you can see, the CRL has considerable ADP needs. These needs are generated by large and variable workload requests combined with limited personnel resources. To satisfy our needs, an ADP system must provide the desired reports from one-time data entry. It must be easy to use and reliable. We are optimistic that such a system can and will be developed in the near future.


STORET
Sample Identification
Field Parameters
Organic
Metals
Inorganic

```
 Use Acnmwdo.wak.mwdo922 on tso017
? list
   1.   1CRLS 05APR DSN=CNMWDO.WAK.MWD0922 /  TS0017                          REV00 T
   2.   -----------------------------STUDY DESCRIPTION--------------------------------
   3.   NPAR NLOG AGENCYID UNLOCKEY STATTYPE SMPLDAY ATLABBY DUEDATE ACCOUNT-NUMBER
   4.     24    3 12MIWID  BRIEMHUR 03444240 15OCT75 17OCT75 03NOV75
   5.   >>>DAIRYLAND POWER,E,J,STONEMAN PLANT,CASSVILLE,WI
   6.   -----------------------------SAMPLE DESCRIPTIONS------------------------------
   7.   LABIDNUM STORETID COLLDAY TIME STATTYPE DEEP T M NO ENDDATE TIME %RU
   8.   767152   DAIRYL   15OCT75        03444240 001  T A 24 16OCT75
   9.   767153   DAIRYL   15OCT75        03444240 003
  10.   767154   BLANK    15OCT75
  11.   >>>767152  >>COOLING WATER
  12.   >>>767154  >>METAL BLANK
  13.   -----------------------------SAMPLE/PARAMETER DATA----------------------------
  14.    EPA-CRL  00010 F  00056 F  00400 F  00500 F  31505 F  31615 F  50060 F      •A
  15.     1975    WATER     FLOW      PH     RESIDUE  TOT COLI FEC COLI CHLORINE     •A
  16.    SAMPLE    TEMP     RATE     FIELD    TOTAL   MPN CONF MPNECMED TOT RESD     •A
  17.    LOG NO.   CENT     GPD       SU      MG/L    /100ML   /100ML    MG/L        •A
  18.   767152  :        :        :        :        :        :        :        : 1S•A
  19.   767153  :        :        :        :        :        :        :        : 2S•A
  20.   767154  :        :        :        :        :        :        :        : 3S•A
  21.              1P       2P       3P       4P       5P       6P       7P          ••A
  22.   .MWD0922   DAIRYLAND POWER,E.J.STONEMAN PLANT,CASSVILLE,WI                   ••A
  23.   ------------------------------------------------------------------------------
  24.    EPA-CRL  00556 OG                                                           •B
  25.     1975    OIL-GRSE                                                           •B
  26.    SAMPLE   FREON-GR                                                           •B
  27.    LOG NO.    MG/L                                                             •B
  28.   767152  :        :        :        :        :        :        :        : 1S•B
  29.   767153  :        :        :        :        :        :        :        : 2S•B
  30.   767154  :        :        :        :        :        :        :        : 3S•B
  31.              8P       9P       10P      11P      12P      13P      14P         ••B
  32.   .MWD0922   DAIRYLAND POWER,E.J.STONEMAN PLANT,CASSVILLE,WI                   ••B
  33.   ------------------------------------------------------------------------------
  34.    EPA-CRL  01067 MW 00927 MW 01034 MW 01042 MW 01045 MW 01055 MW 01092 MW     •C
  35.     1975    NICKEL   MGNSIUM  CHROMIUM  COPPER    IRON   MANGNESE   ZINC       •C
  36.    SAMPLE   NI,TOT   MG,TOT    CR,TOT   CU,TOT   FE,TOT   MN,TOT   ZN,TOT      •C
  37.    LOG NO.   UG/L     MG/L      UG/L     UG/L     UG/L     UG/L     UG/L       •C
  38.   767152  :        :        :        :        :        :        :        : 1S•C
  39.   767153  :        :        :        :        :        :        :        : 2S•C
  40.   767154  :        :        :        :        :        :        :        : 3S•C
  41.              15P      16P      17P      18P      19P      20P      21P         ••C
  42.   .MWD0922   DAIRYLAND POWER,E.J.STONEMAN PLANT,CASSVILLE,WI                   ••C
  43.   ------------------------------------------------------------------------------
  44.    EPA-CRL  01051 MW 01027 MW                                                  •D
  45.     1975     LEAD    CADMIUM                                                   •D
  46.    SAMPLE   PB,TOT    CD,TOT                                                   •D
  47.    LOG NO.   UG/L      UG/L                                                    •D
  48.   767152  :        :        :        :        :        :        :        : 1S•D
  49.   767153  :        :        :        :        :        :        :        : 2S•D
  50.   767154  :        :        :        :        :        :        :        : 3S•D
  51.              22P      23P      24P      25P      26P      27P      28P         ••D
  52.   .MWD0922   DAIRYLAND POWER,E.J.STONEMAN PLANT,CASSVILLE,WI                   ••D
  53.   ------------------------------------------------------------------------------
  54.   1 EPA-CRL  00076 IM 00530 IM 70300 IM 00095 IM 00945 IM 00940 IM 00403 IM    •E
  55.     1975      TURB   RESIDUE  RESIDUE  CNDUCTVY SULFATE  CHLORIDE   LAB        •E
  56.    SAMPLE   TRBIDMTR TOT NFLT DISS-180  AT 25C    SO4      CL       PH         •E
  57.    LOG NO.  HACH FTU   MG/L   C  MG/L  MICROMHO   MG/L     MG/L     SU         •E
  58.   767152  :        :        :        :        :        :        :        : 1S•E
  59.   767153  :        :        :        :        :        :        :        : 2S•E
  60.   767154  :        :        :        :        :        :        :        : 3S•E
  61.              29P      30P      31P      32P      33P      34P      35P         ••E
  62.   .MWD0922   DAIRYLAND POWER,E.J.STONEMAN PLANT,CASSVILLE,WI                   ••E
  63.   ------------------------------------------------------------------------------
? ct
```

Figure 1
Basic Data Form

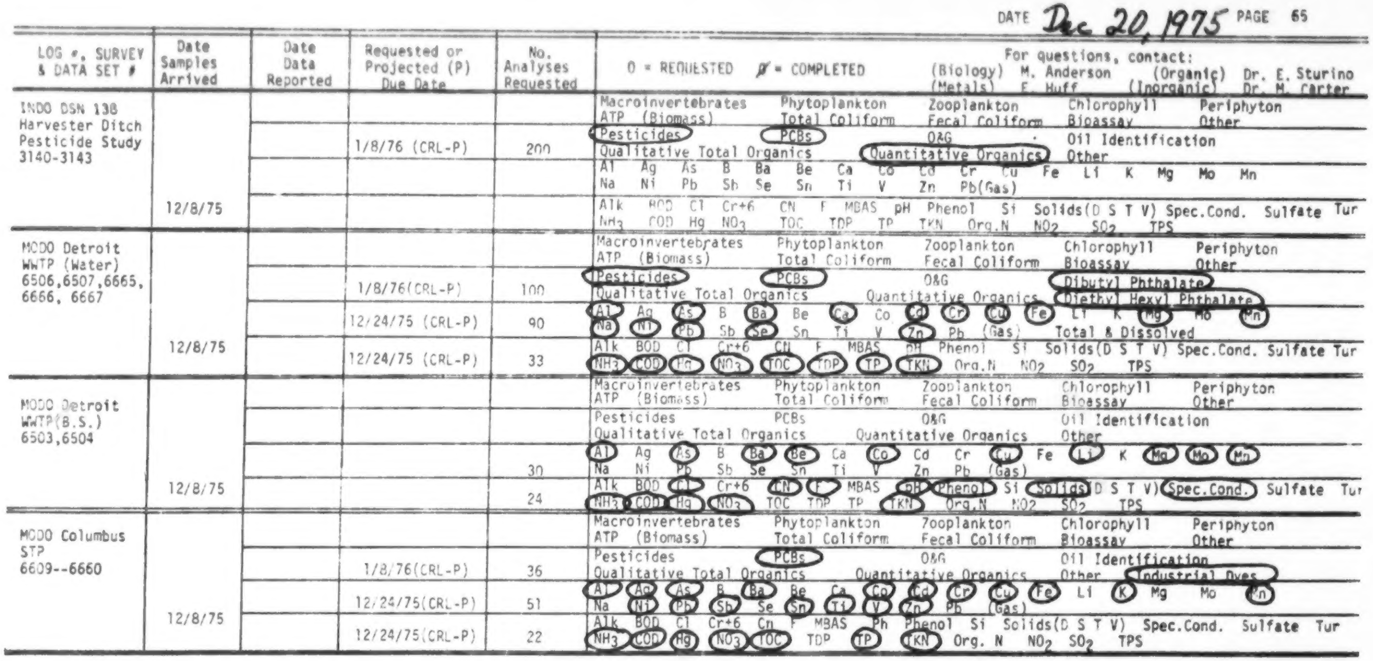


Weekly Report for Inorganic Section

Submitted by Carter Date 10/17/75
Analysis Requested ______ Analysis Completed ______ Backlog X
Position Weeks of Effort ______

Parameter	ILDO	INDO	MODO	MWDO	GLSB	ASB	Other	TOTAL
Alkalinity								
BOD	4							4
Chloride	13		22		159			194
CrIV								
Cyanide								
Fluoride	8							8
MBAS								
pH								
Phenol	15			2				17
Silica	8				200			208
Solids - Dissolved	8							8
Solids - Suspended	8							8
Solids - Total					105			105
Solids - Volatile					105			105
Spec. Cond.								
Sulfate	13				159			172
Turbidity								
Bromide			22					22
Ammonia	17	6	23	3	105		172	319
COD	5	1	8		105			119
Mercury	10				105		344	459
NO_3-NO_2	10	6	23	3			172	214
TOC			17				344	361
TDP					352		172	524
TP	5		10	8	378		172	573
TKN			8	8	282		344	642
PO_4-P							172	172
NO_2(air)						31		31
SO_2(air)						31		31
Solids (air)						44		44
TOTAL	117	13	133	24	2055	106	1892	4340

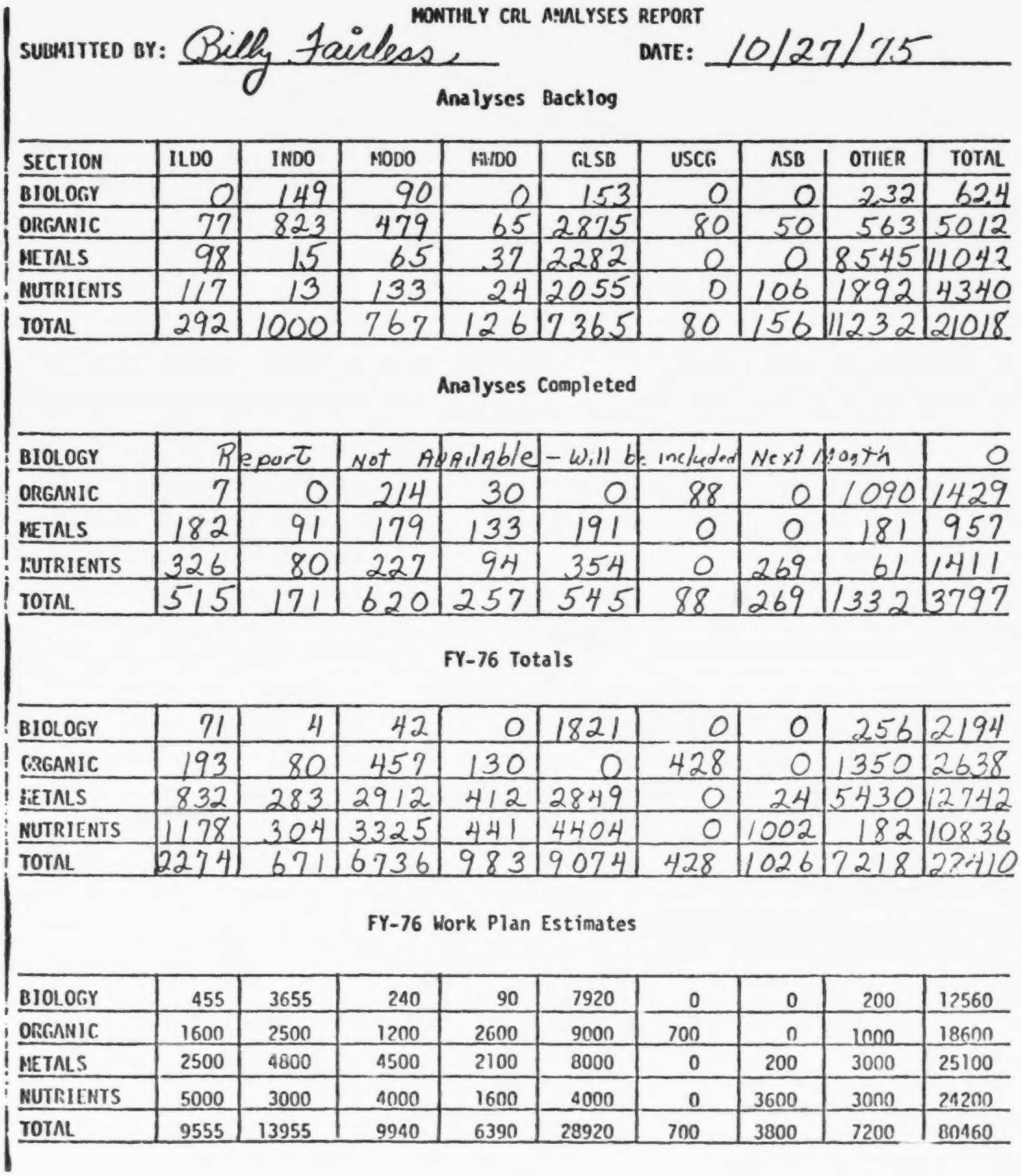
Figure 2
Weekly Report for Inorganic Section



DATE Dec 20, 1975 PAGE 65

LOG #, SURVEY & DATA SET #	Date Samples Arrived	Date Data Reported	Requested or Projected (P) Due Date	No. Analyses Requested	0 = REQUESTED Ø = COMPLETED For questions, contact: (Biology) M. Anderson (Organic) Dr. E. Sturino (Metals) E. Huff (Inorganic) Dr. M. Carter
INDO DSN 138 Harvester Ditch Pesticide Study 3140-3143	12/8/75				Macroinvertebrates Phytoplankton Zooplankton Chlorophyll Periphyton ATP (Biomass) Total Coliform Fecal Coliform Bioassay Other
			1/8/76 (CRL-P)	200	[✓ Pesticides] [✓ PCBs] O&G Oil Identification Qualitative Total Organics [✓ Quantitative Organics] Other
					Al Ag As B Ba Be Ca Co Cd Cr Cu Fe Li K Mg Mo Mn Na Ni Pb Sb Se Sn Ti V Zn Pb(Gas)
					Alk BOD Cl Cr+6 CN F MBAS pH Phenol Si Solids(D S T V) Spec.Cond. Sulfate Tur NH_3 COD Hg NO_3 TOC TDP TP TKN Org.N NO_2 SO_2 TPS
MODO Detroit WWTP (Water) 6506,6507,6665, 6666, 6667	12/8/75				Macroinvertebrates Phytoplankton Zooplankton Chlorophyll Periphyton ATP (Biomass) Total Coliform Fecal Coliform Bioassay Other
			1/8/76(CRL-P)	100	[✓ Pesticides] [✓ PCBs] O&G [✓ Dibutyl Phthalate] Qualitative Total Organics Quantitative Organics [✓ Diethyl Hexyl Phthalate]
			12/24/75 (CRL-P)	90	[✓ Al] Ag [✓ As] B [✓ Ba] Be [✓ Ca] Co [✓ Cd] [✓ Cr] [✓ Cu] [✓ Fe] Li K [✓ Mg] Mo [✓ Mn] [✓ Na] [✓ Ni] [✓ Pb] Sb [✓ Se] Sn Ti V [✓ Zn] Pb (Gas) Total & Dissolved
			12/24/75 (CRL-P)	33	Alk BOD Cl Cr+6 CN F MBAS pH Phenol Si Solids(D S T V) Spec.Cond. Sulfate Tur [✓ NH_3] [✓ COD] [✓ Hg] [✓ NO_3] [✓ TOC] [✓ TDP] [✓ TP] [✓ TKN] Org.N NO_2 SO_2 TPS
MODO Detroit WWTP(B.S.) 6503,6504	12/8/75				Macroinvertebrates Phytoplankton Zooplankton Chlorophyll Periphyton ATP (Biomass) Total Coliform Fecal Coliform Bioassay Other
					Pesticides PCBs O&G Oil Identification Qualitative Total Organics Quantitative Organics Other
				30	[✓ Al] Ag [✓ As] B [✓ Ba] [✓ Be] Ca [✓ Co] Cd Cr [✓ Cu] Fe [✓ Li] K [✓ Mg] [✓ Mo] [✓ Mn] Na Ni Pb Sb Se Sn Ti V Zn Pb (Gas)
				24	Alk BOD [✓ Cl] Cr+6 [✓ CN] [✓ F] MBAS [✓ pH] [✓ Phenol] Si [✓ Solids](D S T V) [✓ Spec.Cond.] Sulfate Tur [✓ NH_3] [✓ COD] [✓ Hg] [✓ NO_3] TOC TDP TP [✓ TKN] Org.N NO2 SO_2 TPS
MODO Columbus STP 6609--6660	12/8/75				Macroinvertebrates Phytoplankton Zooplankton Chlorophyll Periphyton ATP (Biomass) Total Coliform Fecal Coliform Bioassay Other
			1/8/76(CRL-P)	36	Pesticides [✓ PCBs] O&G Oil Identification Qualitative Total Organics Quantitative Organics Other [✓ Industrial Dyes]
			12/24/75(CRL-P)	51	[✓ Al] [✓ Ag] [✓ As] B [✓ Ba] Be Ca [✓ Co] [✓ Cd] [✓ Cr] [✓ Cu] [✓ Fe] Li [✓ K] Mg Mo [✓ Mn] Na [✓ Ni] [✓ Pb] [✓ Sb] Se [✓ Sn] [✓ Ti] [✓ V] [✓ Zn] Pb (Gas)
			12/24/75(CRL-P)	22	Alk BOD Cl Cr+6 Cn F MBAS Ph Phenol Si Solids(D S T V) Spec.Cond. Sulfate Tur [✓ NH_3] [✓ COD] [✓ Hg] [✓ NO_3] [✓ TOC] TDP [✓ TP] [✓ TKN] Org. N NO_2 SO_2 TPS

Figure 3
Central Regional Laboratory Survey Status Report



MONTHLY CRL ANALYSES REPORT

SUBMITTED BY: Billy Fairless DATE: 10/27/75

Analyses Backlog

SECTION	ILDO	INDO	MODO	MWDO	GLSB	USCG	ASB	OTHER	TOTAL
BIOLOGY	0	149	90	0	153	0	0	232	624
ORGANIC	77	823	479	65	2875	80	50	563	5012
METALS	98	15	65	37	2282	0	0	8545	11042
NUTRIENTS	117	13	133	24	2055	0	106	1892	4340
TOTAL	292	1000	767	126	7365	80	156	11232	21018

Analyses Completed

BIOLOGY	R	eport	Not	Available	- Will b	e included	Next	Month	0
ORGANIC	7	0	214	30	0	88	0	1090	1429
METALS	182	91	179	133	191	0	0	181	957
NUTRIENTS	326	80	227	94	354	0	269	61	1411
TOTAL	515	171	620	257	545	88	269	1332	3797

FY-76 Totals

BIOLOGY	71	4	42	0	1821	0	0	256	2194
ORGANIC	193	80	457	130	0	428	0	1350	2638
METALS	832	283	2912	412	2849	0	24	5430	12742
NUTRIENTS	1178	304	3325	441	4404	0	1002	182	10836
TOTAL	2274	671	6736	983	9074	428	1026	7218	28410

FY-76 Work Plan Estimates

BIOLOGY	455	3655	240	90	7920	0	0	200	12560
ORGANIC	1600	2500	1200	2600	9000	700	0	1000	18600
METALS	2500	4800	4500	2100	8000	0	200	3000	25100
NUTRIENTS	5000	3000	4000	1600	4000	0	3600	3000	24200
TOTAL	9555	13955	9940	6390	28920	700	3800	7200	80460

Figure 4
Monthly CRL Analyses Report

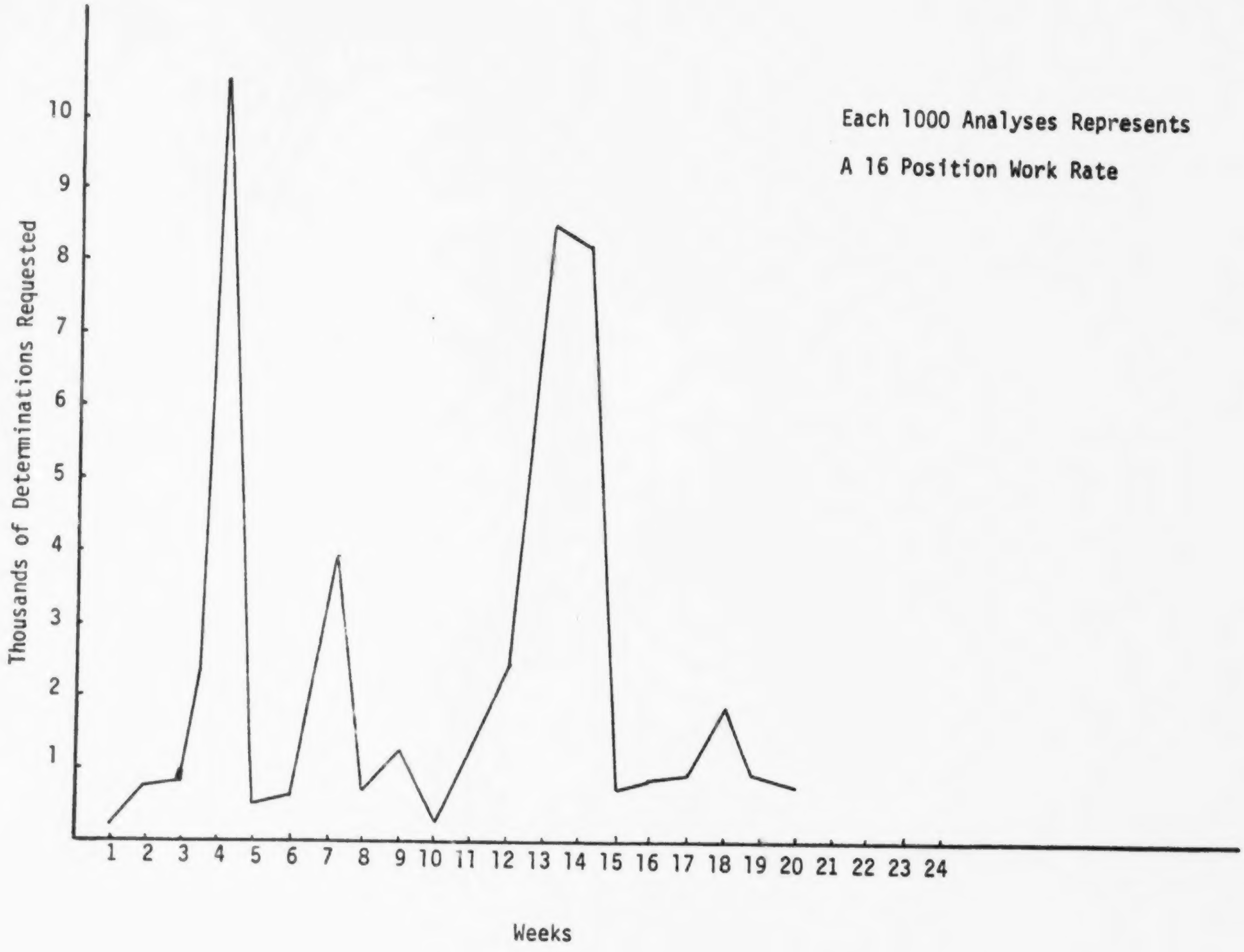
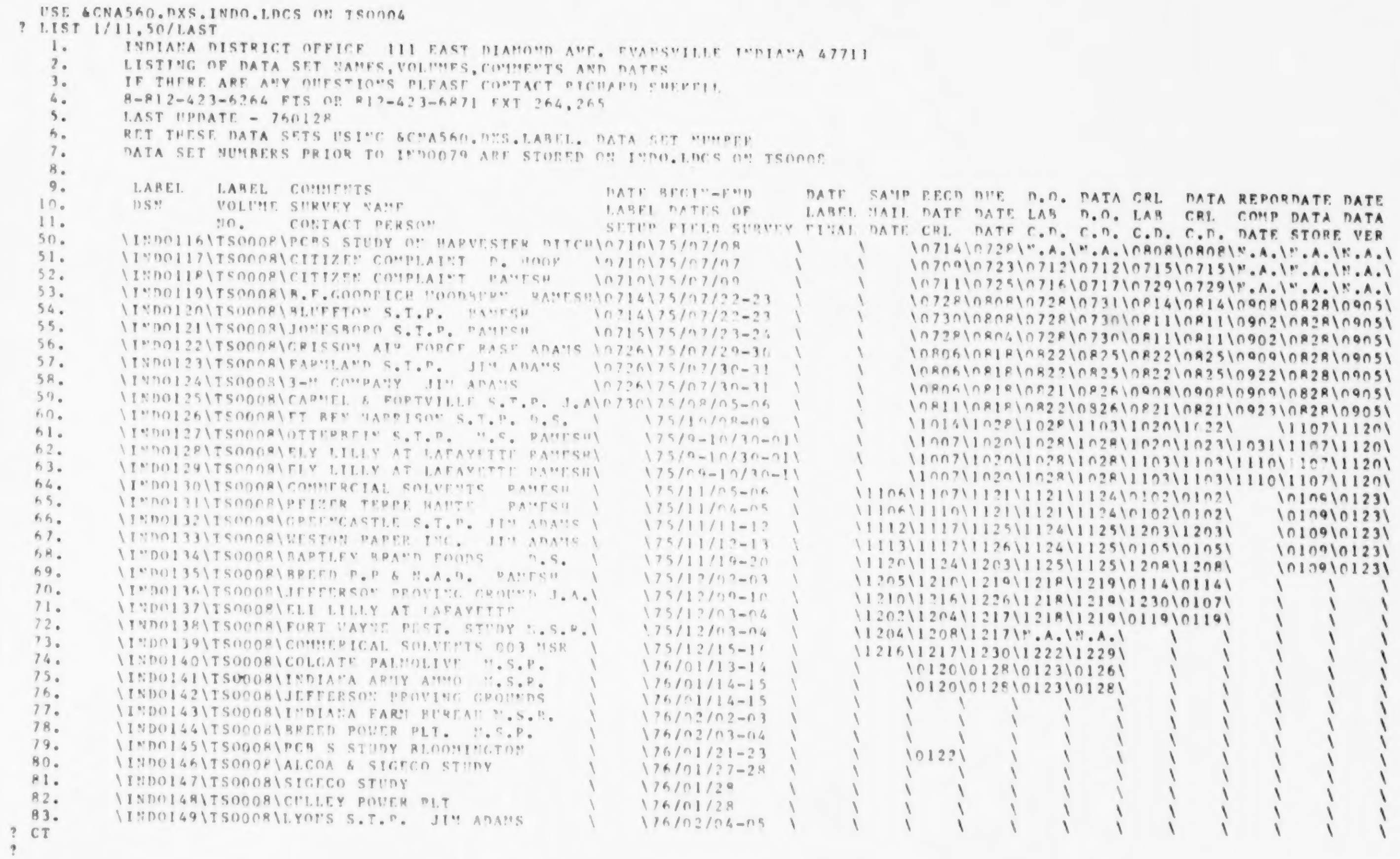


Figure 5
Laboratory Workload As a Function of Time



```
  USE &CNA560.DXS.INDO.LDCS ON TS0004
? LIST 1/11,50/LAST
   1.      INDIANA DISTRICT OFFICE  111 EAST DIAMOND AVE. EVANSVILLE INDIANA 47711
   2.      LISTING OF DATA SET NAMES,VOLUMES,COMMENTS AND DATES
   3.      IF THERE ARE ANY QUESTIONS PLEASE CONTACT RICHARD SHERRILL
   4.      8-812-423-6264 FTS OR 812-423-6871 EXT 264,265
   5.      LAST UPDATE - 760128
   6.      RET THESE DATA SETS USING &CNA560.DXS.LABEL. DATA SET NUMBER
   7.      DATA SET NUMBERS PRIOR TO INDO079 ARE STORED ON INDO.LDCS ON TS0008
   8.
   9.       LABEL   LABEL  COMMENTS                      DATE BEGIN-END     DATE  SAMP RECD DUE  D.O. DATA CRL  DATA REPORDATE DATE
  10.       DSN     VOLUME SURVEY NAME                   LABEL DATES OF     LABEL MAIL DATE DATE LAB  D.O. LAB  CRL  COMP DATA DATA
  11.               NO.    CONTACT PERSON                SETUP FIELD SURVEY FINAL DATE CRL  DATE C.D. C.D. C.D. C.D. DATE STORE VER
  50.      \INDO116\TS0008\PCBS STUDY ON HARVESTER DITCH\0710\75/07/08     \      \    \0714\0728\N.A.\N.A.\0808\0808\N.A.\N.A.\N.A.\
  51.      \INDO117\TS0008\CITIZEN COMPLAINT  D. HOOP   \0710\75/07/07     \      \    \0709\0723\0712\0712\0715\0715\N.A.\N.A.\N.A.\
  52.      \INDO118\TS0008\CITIZEN COMPLAINT  RAMESH    \0710\75/07/09     \      \    \0711\0725\0716\0717\0729\0729\N.A.\N.A.\N.A.\
  53.      \INDO119\TS0008\B.F.GOODRICH WOODBURN  RAMESH\0714\75/07/22-23  \      \    \0728\0808\0728\0731\0814\0814\0908\0828\0905\
  54.      \INDO120\TS0008\BLUFFTON S.T.P.  RAMESH      \0714\75/07/22-23  \      \    \0730\0808\0728\0730\0811\0811\0902\0828\0905\
  55.      \INDO121\TS0008\JONESBORO S.T.P. RAMESH      \0715\75/07/23-24  \      \    \0728\0804\0728\0730\0811\0811\0902\0828\0905\
  56.      \INDO122\TS0008\GRISSOM AIR FORCE BASE ADAMS \0726\75/07/29-30  \      \    \0806\0818\0822\0825\0822\0825\0909\0828\0905\
  57.      \INDO123\TS0008\FARMLAND S.T.P.  JIM ADAMS   \0726\75/07/30-31  \      \    \0806\0818\0822\0825\0822\0825\0922\0828\0905\
  58.      \INDO124\TS0008\3-M COMPANY  JIM ADAMS       \0726\75/07/30-31  \      \    \0806\0818\0821\0826\0908\0908\0909\0828\0905\
  59.      \INDO125\TS0008\CARMEL & FORTVILLE S.T.P. J.A\0730\75/08/05-06  \      \    \0811\0818\0822\0826\0821\0821\0923\0828\0905\
  60.      \INDO126\TS0008\FT BEN HARRISON S.T.P. D.S.  \    \75/10/08-09  \      \    \1014\1028\1028\1103\1020\1022\    \1107\1120\
  61.      \INDO127\TS0008\OTTERBEIN S.T.P.  M.S. RAMESH\    \75/9-10/30-01\      \    \1007\1020\1028\1028\1020\1023\1031\1107\1120\
  62.      \INDO128\TS0008\ELY LILLY AT LAFAYETTE RAMESH\    \75/9-10/30-01\      \    \1007\1020\1028\1028\1103\1103\1110\1107\1120\
  63.      \INDO129\TS0008\ELY LILLY AT LAFAYETTE RAMESH\    \75/09-10/30-1\      \    \1007\1020\1028\1028\1103\1103\1110\1107\1120\
  64.      \INDO130\TS0008\COMMERCIAL SOLVENTS  RAMESH  \    \75/11/05-06  \      \1106\1107\1121\1121\1124\0102\0102\    \0109\0123\
  65.      \INDO131\TS0008\PFIZER TERRE HAUTE   RAMESH  \    \75/11/04-05  \      \1106\1110\1121\1121\1124\0102\0102\    \0109\0123\
  66.      \INDO132\TS0008\GREENCASTLE S.T.P. JIM ADAMS \    \75/11/11-12  \      \1112\1117\1125\1124\1125\1203\1203\    \0109\0123\
  67.      \INDO133\TS0008\WESTON PAPER INC.  JIM ADAMS \    \75/11/12-13  \      \1113\1117\1126\1124\1125\0105\0105\    \0109\0123\
  68.      \INDO134\TS0008\BARTLEY BRAND FOODS    D.S.  \    \75/11/19-20  \      \1120\1124\1203\1125\1125\1208\1208\    \0109\0123\
  69.      \INDO135\TS0008\BREED P.P & N.A.D.  RAMESH   \    \75/12/02-03  \      \1205\1210\1219\1218\1219\0114\0114\    \    \    \
  70.      \INDO136\TS0008\JEFFERSON PROVING GROUND J.A.\    \75/12/09-10  \      \1210\1216\1226\1218\1219\1230\0107\    \    \    \
  71.      \INDO137\TS0008\ELI LILLY AT LAFAYETTE       \    \75/12/03-04  \      \1202\1204\1217\1218\1219\0119\0119\    \    \    \
  72.      \INDO138\TS0008\FORT WAYNE PEST. STUDY M.S.P.\    \75/12/03-04  \      \1204\1208\1217\N.A.\N.A.\    \    \    \    \    \
  73.      \INDO139\TS0008\COMMERICAL SOLVENTS 003 MSR  \    \75/12/15-16  \      \1216\1217\1230\1222\1229\    \    \    \    \    \
  74.      \INDO140\TS0008\COLGATE PALMOLIVE  M.S.P.    \    \76/01/13-14  \      \    \0120\0128\0123\0126\    \    \    \    \    \
  75.      \INDO141\TS0008\INDIANA ARMY AMMO  M.S.P.    \    \76/01/14-15  \      \    \0120\0128\0123\0128\    \    \    \    \    \
  76.      \INDO142\TS0008\JEFFERSON PROVING GROUNDS    \    \76/01/14-15  \      \    \    \    \    \    \    \    \    \    \    \
  77.      \INDO143\TS0008\INDIANA FARM BUREAU M.S.P.   \    \76/02/02-03  \      \    \    \    \    \    \    \    \    \    \    \
  78.      \INDO144\TS0008\BREED POWER PLT.  M.S.P.     \    \76/02/03-04  \      \    \    \    \    \    \    \    \    \    \    \
  79.      \INDO145\TS0008\PCB S STUDY BLOOMINGTON      \    \76/01/21-23  \      \    \0122\    \    \    \    \    \    \    \    \
  80.      \INDO146\TS0008\ALCOA & SIGECO STUDY         \    \76/01/27-28  \      \    \    \    \    \    \    \    \    \    \    \
  81.      \INDO147\TS0008\SIGECO STUDY                 \    \76/01/29     \      \    \    \    \    \    \    \    \    \    \    \
  82.      \INDO148\TS0008\CULLEY POWER PLT             \    \76/01/28     \      \    \    \    \    \    \    \    \    \    \    \
  83.      \INDO149\TS0008\LYONS S.T.P.  JIM ADAMS      \    \76/02/04-05  \      \    \    \    \    \    \    \    \    \    \    \
? CT
?
```

Figure 6
Survey Tracking From Planning to Final Report

DATA COLLECTION AUTOMATION AND LABORATORY DATA MANAGEMENT FOR THE EPA CENTRAL REGIONAL LABORATORY

By Robert A. Dell, Jr.

INTRODUCTION

The Central Regional Laboratory (CRL) is a unit of the Surveillance and Analysis Division within the Chicago-based Region V of the Environmental Protection Agency. The CRL currently performs in excess of 150,000 environmental measurements per year. For the fiscal year 1975, this workload has varied from 100 to 10,000 analyses per week. Thus it has grown to the point where manual handling and reporting of the data have become demonstratively inefficient, and the need for more cost-effective operation with better quality assurance is evident. An example of this need, both the quantity and variability of the workload make advance prediction for lab scheduling difficult. A comprehensive description of the CRL operations and functional requirements for the CRL Laboratory Data Management (LDM) system are described in a companion paper at this workshop.[1] The methodology currently used to automate CRL operations uses both available equipment and personnel, and takes advantage of significant past and ongoing projects within the EPA.

Table 1

Major Components of CRL NOVA Minicomputer System

Model	Description	Cost ($)
8294	840 CPU, 64k of 16 bit words, memory mngmt.	29,300
8206	Power monitor and autorestart	300
8207	Hardware multiply-divide unit	800
8208	Automatic program loader	300
8020	Floating point processor	3,300
4008	Real time clock	300
6003/4019	262K Word Novadisk/control (fixed head)	10,000
4057A/4046	12.472M Word (IBM 2314 type) disc drive/control	21,000
6013/4011	Paper tape reader 400 cps/control	1,600
4030	Mag tape, 9track, 45 ips	9,200
4063	12 line async digital multiplexer	3,700
102A	Centronics printer	6,800
3405	Vadic 1200 baud modem	1,000
4032	Analog/digital converter, 14 bit	4,000
4055J,K,N	16 channel analog multiplexer	1,200
various	Interface cards, cabinets, and DG software	11,400
733	Texas Instruments Corp. console (hard copy) terminals (2)	5,000
ADM-1	Lear-Siegler Corp. (video) terminals (10)	17,000
	Total Cost	$126,000

CURRENT STATUS

The CRL has an automated data acquisition system which was installed in November 1975. Currently two types of measuring equipment are supplying data directly to a Data General Corporation NOVA 840 host computer. The hardware configuration of this system is shown in Table 1 and the documentation is available elsewhere.[2] The two basic prototype data collection programs were delivered with the CRL system for Technicon autoanalyzer and atomic absorption measurements.

At present, the LDM system is in the detailed design stage and essentially allows for the storage, updating, and retrieval of the analyzed data generated by both these programs and manual entry of other measurements. Sample and laboratory supervision is facilitated by incorporating in the LDM system report writing programs. These programs prepare status and short-term periodic reports.

A future aspect envisioned for the LDM system allows for the implementation of an interlaboratory Regional Communications Network (RCN) to further enhance the cost effectiveness of the data base. The RCN would standardize the connection of other instruments and outside agencies to the LDM system.

LAB DATA MANAGEMENT REQUIREMENTS

The LDM system is in response to the needs discussed below.

1. Effective Intra- and Interlaboratory Communications

From study conception through field and laboratory measurement and final reporting, the LDM system must aid communications between the study requester, sample collectors, analysts, and lab supervisors. The format of the report form allows it to be used as an analysis request by the study originator. The same report can be utilized not only as an input mechanism to the STORET system but also for reporting the analysis back to the requester.

2. Timely Report Generation

While the analysis request and study summary are accomplished by the same data set as described above, generation of the workload projections and periodic summaries spanning several studies require considerable manual effort to prepare at present. In particular, the workload listing is essential to program the startup of laboratory processes in an efficient manner.

3. Assurance of In-lab Quality

Included in this area are assurances concerning the parameter measurement precision and accuracy, as well as consistency checks on similar samples, to keep a monitor on the complete analytical procedures.

4. Optimum Utilization of Present Resources

Three areas are identifiable:

A. The EPA Office of Research and Development has cooperated recently with Region V in compiling an Interim Laboratory Data Management System (ILDMS)[3] which supplied several computer programs directly usable in the LDM system.

B. Data collection automation utilizing the on-line programs mentioned previously enters data to the NOVA in computer-readable format. Other instrumentation at the CRL, such as a plasma emission spectrometer and a gas chromatograph, also generates digital data through their associated minicomputer.

C. Standardized Methodology: the NOVA 840 has been used by four laboratories, while a compatible ECLIPSE machine also exists with data management capabilities within the EPA. Remote job entry programs for the OSI and RTP national computer utilities are operational on this machine, and much online program support is available. The foregoing suggests that the LDM system as developed could also be transportable thus allowing other facilities to take advantage of each stage of development. A synchronous data communications protocol and standard LDM implementation languages have also been beneficial.

CONCLUSIONS

At present, the regional laboratories seem to need LDM systems slightly more than automated data collection systems. The effective management of the data, however, requires its reliable entry into computer-readable form. A quality analysis is much easier to obtain with computer-aided detection of peaks and curve fitting routines.

While the EPA-wide effort has been with instrumentation and process control in the past, the LDM system aspects of laboratory automation are ripe for development and can be accomplished with current technology.

REFERENCES

1 Fairless, Billy, "Requirements for the Region V Central Regional Laboratory (CRL) Data Management System" *Proceedings No. 2, ORD ADP Workshop*, 1975.

2 Frazer, J. W. and Barton, G. W., "A Feasibility Study and Functional Design for the Computerized Automation of the Central Regional Laboratory EPA Region V, Chicago," *ASTM STP 578, American Society for Testing and Materials*, 1975, pp. 152-255.

3 EPA Quality Assurance Division, Office of Monitoring Systems, "Development of an Automated Laboratory Management System for the U.S. Environmental Protection Agency," June 1974 and January 1975.

SAMPLE MANAGEMENT PROGRAMS FOR THE LABORATORY AUTOMATION MINICOMPUTER

By Henry S. Ames and George W. Barton, Jr.[*]

In 1973 the Computer Systems and Services Division (CSSD) of EPA-Cincinnati retained the services of a multidisciplinary team of chemists and engineers from Lawrence Livermore Laboratory (LLL) to develop functional specifications for the automation of a number of analytical instruments at the Environmental Monitoring and Support Laboratory (EMSL), Cincinnati.[†] As an outgrowth of that study, LLL was asked to implement the systems specified and also to develop additional specifications and a cost/benefit analysis for the Municipal Environmental Research Laboratory (MERL), Cincinnati, the National Field Investigation Center (NFIC), Cincinnati, and the Central Regional Laboratory, EPA-Region V, Chicago. As a result of these projects, LLL has developed designs for Technicon AutoAnalyzers, several manufacturers' atomic absorption spectrophotometers, the Beckman Total Organic Carbon Analyzer, a Jarrell-Ash Emission Spectrometer, and a Mettler Electronic Balance. These automation designs are now installed in EMSL and Region V on Data General NOVA 840 computer systems.

At this time, LLL is preparing functional specifications for Region III, Annapolis, Maryland. We have also had less formal contact with Regions I and IV, and participated in the Interim Laboratory Data Management Project (ILDMS) sponsored by the EPA Office of Research and Development (ORD). Certain problems of sample and laboratory management have become evident.

For the purpose of this paper, arbitrary distinctions are going to be made among the following:

> *Sample Management*. The tracking of an analytical sample from the time its collection is planned through actual sampling, analytical procedures, and quality assurance to the point of production of a final report on the sample in a form suitable for introduction into an archival data base.

- *Sample Management*. The tracking of an analytical sample from the time its collection is planned through actual sampling, analytical procedures, and quality assurance to the point of production of a final report on the sample in a form suitable for introduction into an archival data base.

- *Laboratory Management*. The additional information necessary for a laboratory manager to plan the work of his laboratory in order to make optimal use of his resources of manpower and instrumentation, or to convincingly document the need for reallocation of resources (e.g., hiring, firing, new instruments, outside contracting).

- *Data Management*. The remaining data reduction functions, including comparison of data with models, investigation of environmental trends, and similar large-scale studies, which require large data bases and powerful processors. We will dismiss these last functions as beyond the scope of the laboratory automation computer installed in the laboratory.

Immediately after LLL produced the preliminary feasibility study and cost/benefit analysis for Region V, we initiated an investigation of what would be needed to provide this sample management capability. A benchmark program was written and demonstrated in DECSYSTEM-10 BASIC. It performed admirably on the PDP-10. Arrangements were made with Data General Corporation to test the program on the computer configuration specified for Region V. Results were exceedingly disappointing. A search of the same simulated data base which ran in less than 1 minute on the PDP-10 took over 20 minutes on the NOVA-840. A number of fixes were considered, but other information became available at about this same time. Although some speed improvements could have been implemented, the original approach was too inflexible to meet potential requirements. The feasibility study for NFIC, Cincinnati, indicated that their sample management system had to satisfy certain legal requirements, and that these requirements might well be placed on regional laboratories too. These requirements included: audit trail, chain of custody, assurance of data integrity, automatic data rejection criteria, and legal defensibility.

---

[*] Speaker

[†] The Lawrence Livermore Laboratory (LLL) is operated by the University of California as a prime contractor to the U.S. Energy Research and Development Administration (ERDA) under contract W-7405-ENG-48. Funds for this project were provided by the U.S. Environmental Protection Agency (EPA) under interagency agreement EPA-IAG-D4-0321 between EPA and ERDA. During the life of this contract both ERDA and EPA have undergone reorganizations. In order to reduce confusion, all organizations and elements will be referred to by their current names.

It became clear that in the long run a totally different approach was needed. Sample management requires at least the following functions in real time or near real time:

- *Sample Log-in*. The online facility that permits survey planners to request analyses, field engineers to enter field data, and laboratory personnel to verify receipt of samples for analysis, including all information necessary to schedule the sample for the needed analysis.
- *Analyst's Workload*. The online facility to permit the operator(s) of the instrument(s) to determine which samples need to be analyzed and to select those which will be analyzed on a given day.
- *Analysis Reports*. The online reports to the analyst of results of analyses, including alerts to the analyst of anomalous conditions (e.g., off-scale, out of range, disagreement of duplicates, unacceptable recovery of spikes, etc.), and summary reports of all the results of the work session. This is installed at Cincinnati and Chicago.
- *Quality Assurance Reports*. Reports to the analyst and his managers of all the relevant quality assurance data including trends, thus permitting the laboratory staff to detect potential problems of precision and accuracy and to take necessary action before, or as soon as, unacceptable results are produced. This is also installed at Cincinnati and Chicago.
- *Consolidated Reports*. Reports to the laboratory manager, the requester of the analysis, and to the national archive system of all pertinent data on a sample or group of samples in a form that requires no further hand transcription with its probability of error.

The LLL approach relies upon separation of real-time functions, instruments which must be serviced on demand, and queries which can wait awhile. A prototype system has been written to investigate the response to be expected in the laboratory environment. It incorporates several of the important features, but in no way can it be considered a product for release. It is, however, a realistic way to investigate the feasibility of this approach to sample management. At the present time, the prototype system, called Sample File Control (SFC), executes on a NOVA 840 at a lower (background) priority than the instrument programs. Communication between instrument programs and SFC is through a buffer area of core accessible to both background and foreground programs. Instrument programs need know only the formats of the SET, GET, LOCATE, etc., calls to access the data. Problems of data base access are handled by the SFC programs. If at some future date it should be desirable to change the format of the data bases, only the SFC need be changed. Instrument programs should require no changes.

Operating on a data base of 25,000 analyses (3 million words), in an environment with ten AutoAnalyzer channels running at the same time, data base queries are answered within 45 seconds. This response time includes the time for a foreground BASIC program to make a request of the SFC, search the data base, and to return data to the BASIC program.

If many instruments (more than 20) are operated simultaneously, response time may become excessive. Response time is largely a function of the time necessary to disk access. Three system changes can be made. The least expensive, which is being investigated now, is to regenerate the MRDOS system so that BASIC and SFC overlays utilize the fixed head disk, with only the data base and BASIC user files on the moving head disk. Thus, the moving head disk controller would be able to search the data files with a minimum of head motion introduced by non-SFC activities.

A moderately expensive enhancement to the system would be to increase the size of data buffers by the purchase of additional core. The number of disk accesses would be reduced in inverse proportion to the buffer size, and the search speeded up comparably. The most expensive enhancement would require the purchase of an additional separate processor which would have SFC or one of the commercial data base management systems as its highest priority job for communication with the archival systems, report generation, management queries, and so forth.

We have concluded that an acceptable SFC must not require various undesirable compromises such as requests for the analyst's workload files the previous night, production of consolidated reports overnight, and manager's status reports only as of the end of the previous day. It must be flexible and must not require the modification of user programs in order to include new analytes or additional information such as audit trails.

The present approach described here is flexible and open-ended; it admits a number of enhancements as analytical requirements change and as analytical load increases. The investigation of sample file management alternatives is nearly complete, and a valuable and flexible sample management system could be installed by mid-1977.

SUMMARY OF DISCUSSION PERIOD - PANEL II

The presentations concerned with laboratory data management (LDM) generated a number of questions.

Need for Automated LDM

It was suggested that a sample load of 10,000 to 100,000 analyses per year was sufficient to make the use of an automated LDM system feasible; however, it was pointed out that this really depends on the resources available to a laboratory. Clearly, 1,000 analyses per year would be sufficient if the LDM program were available cheaply enough.

Sources of Assistance

Two sources of assistance in LDM were identified within EPA. The Management Information and Data Systems Division in Washington, D.C., has basic ordering agreements with several contractors who will assist in the development of such requirements specifications and documentation. The Environmental Monitoring and Support Laboratory in Cincinnati will assist EPA monitoring laboratories in the implementation of laboratory automation/LDM systems. There are several commercial minicomputer-based data management systems available. These are relatively new but are being considered in feasibility studies for LDM systems. Finally, it was pointed out that the economies of scale imply that EPA monitoring activities should be consolidated into a small number of highly automated, very efficient laboratories. However, several of those present expressed serious reservations about the wisdom of this approach.

Specification Development

There was considerable discussion about the development of specifications for LDM systems. It was pointed out that simple flow chart-level specifications are usually too vague and general and that implementations based on them alone lead to problems. Detailed functional descriptions are required to avoid acquiring an unwanted system. There was discussion of the justification for the response time requirements in the new Region V LDM specification. Finally, the delays involved in implementing LDM systems were discussed. These were attributed to vague, inadequate specifications, higher priority instrument automation, underestimated required resources, and limited resources.

SHAVES and NASN

There was a great deal of interest in two existing LDM systems that use the traditional data processing center approach. These systems are the EPA, Corvallis, "SHAVES" system and the system developed for the National Air Surveillance Network. Points covered included development costs, use costs, turnaround time, implementation time, personnel requirements, programing language interfaces, and batch versus interactive operations.

Standardization

The desirability of standardization was mentioned several times. For certain classes of laboratories (e.g., environmental monitoring), standardization is highly desirable and could result in significant cost savings. It was pointed out that differences in hardware and operational methods work against standardization.

Languages

There was some discussion of the merits of programing in assembler and high level languages. Several participants agreed that the number of lines of fully debugged code produced per unit time by an experienced programer was independent of language. It also was pointed out that one line of high-level code usually accomplishes as much as many lines of machine language code.

OSI System and STORET

It was asserted that the interim LDM system (OSI-based) will complement minicomputer-based systems, especially in regard to interfacing with STORET. The consensus was that quality assurance concepts should be applied to environmental data before its transfer to archival storage (e.g., STORET).

THE STATE OF DATA ANALYSIS SOFTWARE IN THE ENVIRONMENTAL PROTECTION AGENCY

By Gene R. Lowrimore

Data analysts have yet to realize the full benefits promised by the availability of large-scale computer power. We keep looking forward to the day when we can concentrate on the analysis of the data, that is, the day in which the hardware-software machine will enable us to do the analyses we want to do with only a reasonable amount of nondata analysis effort. This paper briefly discusses the current situation within EPA concerning data analysis software and outlines what kind of software we might develop to support the data analysis effort more fully.

First of all, who is a data analyst? For purposes of this discussion, the following definition will be used:

> Data Analyst–One who analyzes, for some purpose or other, data. For the most part, he or she is assumed to know what operations should be performed on the data in the process of analyzing it.

The software tools which a data analyst has at his or her disposal are of three kinds: (1) subroutines, which, of course, require the writing of a main program before anything useful can be done with them; (2) stand-alone programs, which do not require additional programing but require that the data be input at least once for each analysis desired; and (3) integrated packages, which allow many analyses to be performed on the data once it has been entered. The following list is representative of the kinds of software tools available in EPA.

Of the collections of subroutines, STATPACK and SPSS are vendor products (IBM no longer supports SSP). ARL Linear Algebra Library and Box-Jenkins Time Series Analysis were developed by some data analysts for their own particular purposes and furnished to us to use at our own risk. IMSL is generally considered to be the best general purpose subroutine collection available.

Subroutine Collections	Stand-Alone Programs	Integrated Packages
Univac STATPACK	BMD-P Series	Statistical Analysis Systems (SAS)
Scientific Subroutine Package (SSP)	MANOVA	Statistical Package for Social Sciences (SPSS)
ARL Linear Algebra Library	Multivariate General Linear Hypothesis (MGLM)	
Box-Jenkins Time Series Analysis		OMNITAB
International Mathematical & Statistical Library (IMSL)	Linear Categorical Analysis (LINCAT)	STATJOB

Among the stand-alone programs, the BMD-P series is probably best known to most data analysts. It is a series of 80 or more programs furnished by the UCLA Computer Center. These programs were developed for the IBM 360/370 series computers. They have been converted to the Univac 1100 series computers and are distributed in this form by the University of Maryland. The other three programs in the list are supplied by the University of North Carolina at Chapel Hill. MANOVA and MGLM are quite useful for performing extensive multivariate analyses. LINCAT analyzes contingency tables by exploiting the analogy with the analysis of variance techniques.

From a data analyst's point of view, SAS (developed by North Carolina State University) is by far the best of the integrated packages available. The strength of SAS is its data handling capability and the ease with which the data analyst can invoke any procedure within its repertoire. SPSS is supplied by the National Opinion Research Center (NORC) and is the strongest of the integrated packages for making contingency tables and performing descriptive statistics. OMNITAB was developed by the National Bureau of Standards and is useful primarily as a data analysis tool for limited analysis on small data sets. STATJOB was developed by the University of Wisconsin and appears to be an enhancement of SPSS. The 7 inches of documentation accompanying the package is a major obstacle to STATJOB's use.

These tools are typically those with which an EPA data analyst must work. Because the software has not been adequately designed for general use, a programer is usually assigned to assist the data analyst in carrying out the analysis. When a programer is unavailable, the data analyst must perform that function. Since this kind of programing is unexciting, it is likely to be greeted with something less than enthusiasm. This lack of enthusiam frequently leads the data analyst to perform the programing function even when not absolutely necessary.

Just as frequently, the data analyst gets so involved in programing that he or she ceases to be an effective data analyst.

Some of the packages and programs available to the data analyst are excellent. For instance, the availability of SAS has dramatically cut the overall time required to complete an analysis of data. SAS is a prime example of what good software can do to improve data analysis, but we should not stop with SAS, SPSS, or any other particular package. Much data analysis needs to be done, for which software is not available. The functions in data analysis done manually are very expensive in terms of money, time, and accuracy.

EPA needs to develop, or cause to be developed, data analysis packages which will greatly reduce the need for assigning programers to assist data analysts, reduce the overall time and expense of doing data analysis, and increase the scope of the analysis that the data analyst can easily do. In order to accomplish these ends, the software should:

1. *Handle Large Data Sets Effectively.* None of the available packages has this capability. For instance, SAS stores all data in double precision. SAS also reads the data set several times unnecessarily.

2. *Manage Analysis Results.* One of the requirements facing EPA researchers is that the analyses included in published reports should be exactly reproducible by different programs. Under the Freedom of Information Act, industry is requesting EPA data and subjecting it to their own analyses. The system should keep up with the results and the observations used in each analysis.

3. *Make Good Use of Plotting Capabilities.* Publication quality plots or printer plots should be as easy to generate as any other statistical procedure. Presently, a program must be written in order to generate a publication quality plot. This situation is unreasonable and unacceptable.

4. *Operate Effectively in the Interactive Mode.* Abbreviated output for a procedure should be sent to the interactive terminal; the complete computer output should be saved and sent at the end of the session where the data analyst states. This software should lead the data analyst through the session to whatever extent necessary.

5. *Use Analysis Techniques Which Exploit Computer Capability.* Procedures are needed for exploratory analysis of large data sets. The capability to use empirical sampling techniques to test hypotheses needs to be provided. Least squares procedures that do not require the formation of the normal equations should be used. SAS would not have required double precision data representation if this had been done.

6. *Allow the Analyst to Estimate Power of a Test of Hypotheses.* The power of the hypotheses test being used by the data analyst is rarely calculated, primarily because the computation is difficult.

7. *Incorporate User Procedures.* It should be recognized at the onset that everything a data analyst might want to do cannot be anticipated. Therefore, linking an analyst's pet procedure with the data handling capabilities of the system should be made as simple as possible.

John Tukey was discussing these same problems in 1963.[1] The question obviously arises, "Why haven't we gotten more of this capability in the ensuing 12 years?" The answer has three parts:

- A false definition of scientific programing has been promulgated which says that very little input and output are involved and much calculation is required

- Data analysts often have not let their thinking about the process of analyzing data be influenced by the presence and power of the computers. Consequently, they have not provided the necessary analysis support nor have they been demonstrative enough in demanding the right kind of ADP support

- Computer programing has been a very unreliable enterprise. Only in the last few years has some structure been brought to the programing process. This structure will give us confidence that we can successfully develop more comprehensive systems.

As ADP professionals, the challenge to us is to turn the situation around and really accomplish something in the field of data analysis software.

REFERENCE

1 Tukey, John W., "The Inevitable Collision between Computation and Data Analysis," *Proceedings IBM Scientific Computing Symposium Statistics.* IBM Data Processing Division, White Plains, New York, 1963, pp. 141-152.

NATIONAL COMPUTER CENTER (NCC) SCIENTIFIC SOFTWARE SUPPORT - PAST, PRESENT, AND FUTURE

By M. Johnson

Traditionally, the computer center at Research Triangle Park has obtained and maintained scientific software packages of general utility for its user community. Until 5 years ago, there was an IBM 1130 computer serving about 25 local users in what was then the National Air Pollution Control Administration. Nearly all processing was of a scientific nature using FORTRAN, the 1130 statistical and mathematical package, APL, and SPSS. A great deal of CALCOMP plotting was also done.

During the next 4 years after the installation of the IBM 360/50, the user community expanded and applications greatly diversified. Statistical packages such as BMD, SPSS, and SAS were made available, as well as the IBM Scientific Subroutine Package. The interactive TSL library was made accessible under TSO. Thus, scientific software was implemented and maintained, but no central consulting or training support was available. Users tended to help each other with problems; trial and error was the methodology. Also, as is frequently the case when no designated user support staff exists, the last system programer "touching" a particular package tended to become responsible for it by default. Any time spent diagnosing user problems was at the expense of other regularly assigned systems tasks.

As the 360/50 became saturated about a year after its installation, time was bought from the neighboring university computer center. This provided access to a wider range of statistical packages as well as to APL. The universities also offered several short courses in such packages as SAS and SPSS, at both beginning and advanced levels, which were available to the EPA user community.

Who made up the user community at that time? Primarily it was still local, consisting of the Office of Air Quality Programs and the RTP National Environmental Research Center, but had grown to well over 100 users. Regional offices began retrieving data from the SAROAD data files. However, scientific applications still represented a large portion of the job mix.

Shortly before the installation of the Univac 1110 in the fall of 1973, a user services function was established. Unfortunately, this coincided with the departure of the two DSD staff members who had statistical backgrounds and experience with scientific software. These vacancies could not be refilled, so scientific users still had to rely primarily on each other for debugging assistance. Naturally, a situation developed that once a routine was made to work and the user became familiar with a certain software package, there was great reluctance to explore other possibly more expedient alternatives.

Procurement of scientific software for the Univac system has been rather haphazard, although the computer center has attempted to obtain and implement available software in response to user's requests. SAS and TSL were converted as part of the Univac conversion contract. STAT PACK and MATH PACK are Univac-supplied routines directly callable from FORTRAN. OMNITAB was obtained from the National Bureau of Standards, STATJOB received from the University of Wisconsin, and APL acquired from the University of Maryland. The University of Maryland now has the BMD-P series available for Univac, and this package has been ordered for NCC. All the standard CALCOMP plotting capabilities are available and TEKTRONIX interactive graphics have been implemented. A Univac version of SPSS was one of the first packages to be installed, but installation was essentially where central support stopped. Consultation and debugging assistance has been severely limited and training virtually nonexistent in the efficient and expedient use of the software.

Recently, a scientific software manual has been developed by SAI under contract for Elijah Poole. SAI surveys Agency-wide scientific software and provides descriptions and sample runstreams of all available packages. We now have a central source of current information which can be expanded and updated as our resources improve. Elijah Poole has also coordinated three regional training sessions to introduce the manual and software resources to the EPA scientific community.

Now as a part of MIDSD under the direction of Willis Greenstreet, we are on the threshold of change and the future looks bright. We are now charged with the responsibility of becoming an Agency-wide computing resource serving several hundred users and have a new name, EPA National Computer Center. Not only is it necessary to upgrade the computer hardware but also to improve supporting services. A modification to the existing systems programing contract with ISSI will provide a highly experienced staff dedicated to the support of user services functions. The responsibility of enhanced scientific software support has been clearly defined and appropriate staffing is being procured. This support will include evaluation, implementation, maintenance, documentation, training, and consultation—all specific to the needs of the NCC user community. The Scientific Software Committee will be resurrected and made a viable channel of information exchange within the scientific community. Attention has been brought to bear on the inadequacies of the converted version of SAS and on the question of what software, currently unavailable, should be provided. New software requirements will be evaluated in a reasonable and orderly manner and, once procured, adequate support will be provided to assure efficient utilization.

EXPLOITATION OF EPA'S ADP RESOURCES: OPTIMAL OR MINIMAL?

By John J. Hart

The traditional approach to the provision of ADP support to various functional requirements in Government and industry has been based on centralization of hardware, software, and systems analysis/programing resources. Because of economies of scale and requirements for highly specialized technical skills, this concept has been both necessary and desirable. Generally, individual divisions and branches within EPA research laboratories cannot afford to employ a central staff of analysts and programers with the varied technical skills necessary to effectively support the diversified data processing and analysis functions associated with today's research and development problems. Likewise, the costs of complex, sophisticated, and comprehensive hardware/software systems prohibit the use of local computer installations. Although for all practical purposes EPA's centralized hardware, software, and personnel resources are providing competent and useful support services, it is suggested that the Agency has not yet fully exploited the total capabilities available and inherent in the sophisticated computer systems at the National Computer Center (Univac 1110) and Optimum Systems, Inc. (IBM 370). This paper will review several factors which affect the utilization of these capabilities and will suggest opportunities for improvement.

The Agency's missions include the identification of pollutants, overall assessment of environmental quality, development of strategies and techniques for control and abatement, and implementation of continuous monitoring functions and mechanisms. The scope of these missions and the quantity of physical, biological, and chemical parameters which must be measured, analyzed, and interpreted, obviously confront the Agency with an enormous information and data explosion. Consider the possible outcome if EPA were limited to the technology available in the 1950's. An enormous number of people would be performing statistical computations with electromechanical calculators and slide rules, the overall productivity would be low, and the error rates would be extremely high. The ability to implement and effectively use sophisticated modeling and simulation would also be severely restricted.

Fortunately, in 1976, the Agency has the sophisticated and comprehensive ADP resources to solve the complex analytical and processing requirements associated with its research and development missions. In addition to numerous dedicated laboratory minicomputers used to support analytical instrumentation, the Univac 1110 at NCC and the IBM 370 at OSI provide extremely fast computational and processing capabilities, substantial mass data storage facilities, and extensive libraries of canned scientific programs to support statistical analysis, modeling, and simulation. High-level programing language processors (FORTRAN, COBOL) and software systems to efficiently support data entry, editing, and data base file structuring (Wylbur, IRS, System 2000) are also available. Through the existing time-shared low-speed terminals and multiplexed communications facilities, scientists, programers, and data clerks have immediate access to the large-scale computers for program implementation and execution, and for performance of varied data handling functions (e.g., entry, editing, and retrievals).

With all of these resources and capabilities, it would be natural to assume that their application to the Agency's missions are cost effective, efficient, and sufficiently comprehensive in scope. It is suggested that these conditions do not accurately describe current conditions. For example, let us examine several characteristics pertaining to the attitudes and involvement of management and the scientific community. Frequently we hear the following concerns expressed by management:

- Extremely large ADP expenditures
- Lack of ADP planning
- Fragmented and nonstandardized ADP resources and approaches to supporting the Agency's missions
- Complexity of issues regarding what resources are required for specific applications
- Complex technology requiring specialized knowledge and training

- Communications problems in interfacing with ADP professionals

- Lack of adequate and competent assessment of the cost benefits obtainable through use of ADP technology and resources.

In contrast, the scientific community is usually too busy promulgating the Agency's technical missions to become intimately involved with the proper planning and application of ADP resources. In fact, the scientific community can be classified into three distinct groups:

1. A group which is significantly indifferent to ADP technology and issues. They appear to execute their technical data analysis responsibilities without computers.

2. A group which acknowledges the need for use of automation techniques and resources for selective problems. Typically, they depend on central ADP professional staffs for selective applications development.

3. The last group is very active in the application of ADP resources in that ADP is integrated into their technical line responsibilities relating to statistical analysis, modeling and simulation, and engineering design. In many cases, these people have taken the initiative to learn computer programing and have developed excellent skills equal to, and greater than, many ADP professionals. To them, ADP resources become effective tools for analysis, design, and simulation functions.

The previously mentioned areas of management concerns and interests and involvement of scientific personnel have a direct impact on the extent to which ADP technology and resources satisfy the Agency missions. The impacts are manifested in the following ways:

- Low numbers of ADP users among the scientific community

- Persistent use of manual methods for data analysis

- Low analytical productivity

- High error rates

- Redundancy in ADP systems development

- Arms-length relationships with the ADP professional and difficulties in communications

- Disproportionate expenditures of ADP funds (e.g., administrative vs. research and development).

Can these conditions be changed? Can ADP resources be more effectively employed? Is it possible to increase the use of ADP in support of the technical missions of the Agency? The answer to these questions is affirmative. However, the burden of such accomplishments rests with the ADP professionals and their respective management. First of all, the ADP professional will have to take the initiative to break down the communication barrier by the following means:

- Simplify the language used in communicating with prospective users concerning the design and implementation of new applications; deemphasize the ADP technical jargon

- Develop an increased awareness of the functional aspects and utility of the user's proposed application in the user's environment

- Determine, describe, and emphasize the cost benefits to be derived from the proposed application

- Determine and formulate the disciplines and procedures required to make the application successful and effective.

A second requirement for increasing the effective application of ADP technology to the Agency's missions will be to develop strategies and programs which enlarge the ADP user population representing the scientific community. The following are some possibilities:

- Develop training seminars which demonstrate and describe typical types of computer applications using information from existing systems

- Develop an introductory training course on standard general purpose statistical packages (e.g., OMNITAB, SAS, BMD); such a course would present an overview of the unique capabilities of each package and typical problem applications

- Develop an introductory training course on the standard graphics packages and their application to typical Agency problems; this course would be followed by additional in-depth instructional courses on specific packages (i.e., Calcomp, IPP, Tektronix)
- Develop an introductory ADP concepts course which enumerates and describes the Agency's ADP resources and facilities (e.g., OSI, RTP, Univac, IBM 370, Wylbur, other buzzwords).

The purpose of these suggested courses is to indoctrinate the scientific community on the available resources and typical applications. Existing courses, such as FORTRAN programing, System 2000, and Wylbur text editing, provide the detailed training required for use of unique individual systems.

Additional effort is also recommended for expanding the use of existing statistical packages and for reducing the complexity of using several packages. There are four specific recommendations to be made for accomplishing this objective.

1. All statistical packages at OSI and NCC should be reviewed and tested. Each package should be tested to determine overall functional capabilities, limitations, and restrictions for use.

2. A cross-reference directory should be developed to help a scientific user select the package best suited to his/her requirements. This cross-reference directory should briefly describe the functional capabilities of each package and identify all packages which solve common problems (e.g., analysis of variance).

3. Simplified written procedures should be developed for use of selected statistical packages to solve common and frequent types of problems (e.g., simple regression). It has been suggested by several laboratory scientists that a cookbook procedure for SAS and OMNITAB would be useful for selected problems.

4. Considering the redundancy of statistical software packages installed on Agency computer systems, a comprehensive review of all existing packages followed by development of a standard package for the Agency may prove useful. At present, the scientist is confronted with the task of reviewing literature on multiple statistical packages, each of which was designed for a unique scientific discipline (e.g., behavioral science, medical) and has unique characteristics and limitations.

SCIENTIST, BIOMETRICIAN, ADP INTERFACE

By Neal Goldberg

To look at strengths and weaknesses in the scientific analysis of data in EPA, we must look well beyond the realm of automatic data processing (ADP). A general picture must include consideration of three disciplines:

- Scientist
- Biometrician-statistician
- Computer specialist.

There must be a viable working relationship among these three persons, or groups of persons, to ensure proper completion of a project. Subsequently, there must be a shared understanding of major principles applied by all involved with the study.

Generalizing, the scientist may be a chemist, biologist, environmentalist, and so forth. In the scientific scheme, the scientist is the person who defines a problem, accumulates data, and presents his or her solution. The scientist must rely upon the biometrician who is concerned with the proper design (e.g., replication, sampling techniques, etc.) and interpretation of data. Essentially, the biometrician will instruct the scientist in the proper statistical procedures to find what he is looking for and explain what he has actually found. In any but the most basic experiment, an unwieldy amount of data is usually accumulated. In most cases, the biometrician needs to call upon the computer specialist for proper organization and application of his/her requirements in the interpretation of data.

Within the laboratory, we see too many experiments which are not "designed" until "after the fact." Most often, this error is caused by failure to comprehend proper mathematical/statistical techniques for sampling and hypothesis validation. It is possible that the effects of this misinterpretation could cause irreversible damage if left uncorrected.

This error referred to is evidenced in many forms, the most notable being that of lost time and money. The most serious, however, involves the loss of credibility. One example, is cited below.

A senior member of the scientific staff at a research laboratory undertook an experiment to study the effect of a potentially toxic substance using a suitable biological indicator organism. At the conclusion of the 6-month sampling period, he came to the slow realization that he did not understand how to validate his findings mathematically. In an attempt to find a reasonable solution, the ADP operation was requested to provide a variety of statistical analyses. All software was available from existing proprietary packages. Upon completion of these analyses, no useful information was found. All data were then rerun utilizing logarithmic transformation. After 5 months of attempting to find a solution, a stop was put to the processing of these data. At this time, over 100 jobs had been run, requiring the punching of about 3,000 cards, more than 60 graphs were produced, and more than 80 online data sets were maintained. Total direct ADP costs incurred were around $2,500.

The principal investigator and others within the lab began seriously to seek biometric services. After consultation with members of the Department of Experimental Statistics at a major university, all were satisfied that a reasonable solution had been found. In one day, five jobs were run and proper interpretation of the data provided. Total cost for direct ADP services was $25 with only 560 cards required to be punched.

A loss of Agency credibility would reach far beyond the loss of time and money cited previously. It would strike directly at the justification for our organized existence (i.e., protection of the environment).

A large portion of EPA effort is enforcement oriented. In order to remain an effective entity, we must be able to provide constructive support. The data we produce must serve this end. Therefore, it is essential that our methods (i.e., experimental design and statistical/mathematical reduction) must be defensible as well. Opponents of some of EPA's policies will expend vast resources in an attempt to invalidate the Agency's findings and weaken its regulatory ability. We cannot afford to allow a defeat based upon a technicality in experimental procedure.

Granted, these problems do not exist at all facilities, but they have reached a critical point in some areas. Currently at most Environmental Research Laboratories, it is recognized that better utilization of existing data is required. There are cases where new experimentation is in a holding stage pending adequate biometric analysis. Also, interpretation and publication of some existing data are being withheld until procedural methods are suitably defined.

Where all three personnel resources are available with effective lines of communication, an efficient scientific process is more than likely to be found. In cases where this effective process is not found, it can be assumed that the parties involved will seek to nullify the predicament. It is, therefore, necessary to address those with the authority to act. Scientific problems of the previously discussed nature are such that they are difficult, if not impossible, to communicate via the telephone. Currently there are means for clarifying both scientific and ADP problems.

EPA should attempt to equip those in the field with the proper support for biometric/statistical functions. Carrying this one step further, an investigation should be undertaken to study the feasibility of providing a staff to travel between laboratories, providing services as required. This probably would give rise to a significant improvement in data handling and enhance the atmosphere for more interlaboratory communication. Such solution would have the least impact upon the personnel shortage.

There is general agreement between the scientific investigator and the computer specialist that there is a need for an intermediary in the scientific process; yet, it is extremely difficult to acquire one. Neither the contracts office nor the personnel office have been able to provide any direction toward this goal.

The EPA scientific software study shows that sufficient mathematical/statistical analyses are available to meet most requirements. This wide spectrum of proprietary software is of little value, however, when there are so few available who fully understand its utility. In this context, it must be recognized that ADP services cannot be used as a tool for arbitrarily attempting to solve problems. Like all others, this resource is not limitless.

To minimize problems, project managers and ADP coordinators must be sure that they are getting the most out of a project. Automatic data processing cannot be scientifically effective as an entity. It is only one part of a whole system and, as such, cannot function until all the pieces are in place.

STATISTICAL DIFFERENCES BETWEEN RETROSPECTIVE AND PROSPECTIVE STUDIES

By Dr. R. R. Kinnison

The U.S. Environmental Protection Agency is currently collecting and archiving massive quantities of environmental data. This collection is intended to provide rapid access to answers to questions that may arise in the future. With this rapid access, it will not be necessary, in most instances, to initiate a project to collect raw data for an adequate analysis of a situation.

In statistical terms, analysis of existing data is performed in a retrospective study. Development of a data base to answer an existing question is performed in a prospective study. Statisticians know that the generally used statistical data analysis techniques yield biased answers when applied in retrospective studies. The commonly used statistical data analysis techniques were developed for prospective studies, and they assume many characteristics of the data base, some of which cannot be met by a retrospective data base. The usual resulting bias is in the direction of finding significant effects when, in fact, none exist.

Of course, the concept of a retrospective study is not new; it was developed to a high degree of sophistication by medical epidemiologists studying chronic diseases in humans. In such instances, a prospective study would require following a sample of humans for their entire lifespan of about 70 years. Such a study obviously would be expensive, and the time delay between question and answer would be unacceptable. In addition, elaborate precautions are necessary just to keep track of the fate and the location of the elements of the sample. Naturally, techniques have been developed to avoid a prospective study in such situations. These techniques come under the general statistical category of "Observational Studies," and, in general, they rely on retrospective data.

The statistical analysis tools applicable to observational studies are distinct from those applicable to prospective studies. They are not as highly developed as the more commonly employed statistical tools. There is also a substantial amount of current research effort being devoted to this type of statistical analysis. An excellent review article, "The Design and Analysis of the Observational Study–A Review," by Sonja M. McKinlay, was published in the *Journal of the American Statistical Association*, September 1975 (Volume 70, Number 351). Those characteristics of observational studies that specifically define a retrospective analysis follow. The "treatment," which determines the groups for comparison, is the statistical effect, and the observed response is the hypothesized cause. Thus, we look at the past smog exposure in people that now have lung disease, rather than determine what will happen in currently healthy people exposed to known degrees of smog.

Those assumptions of "regular" statistics that are violated in retrospective analysis are that: (1) the replications of the experiment are made under similar and known conditions, (2) the replications are mutually independent, and (3) the uncontrolled variation is due only to random fluctuations. In retrospective studies, the conditions of the experiment are random and the effects are known. Thus, the treatments cannot be preassigned in a random manner. Note that a retrospective study invariably will exclude all those subjects exposed to the treatment, who did not develop the effect. Without these nonresponders, common statistical techniques are invalid. Two other properties are desirable in retrospective studies; techniques that permit (1) the capability to measure and to manipulate systematic error in the absence of randomization, and (2) methods to evaluate evidence from a variety of sources, each of which may have unknown and different characteristics.

The EPA computerized information systems traditionally have been used to collect data, retrospectively evaluate that data to suggest questions, then used the same data to answer those questions using techniques applicable only to prospective studies. The data are good, but the answers are not. There is a statistical science available that will provide good answers. Our current data screening efforts to find good questions are valid, but the finding of a possible effect is not synonymous with the firm and statistically valid establishment of the existence of that effect. We have taken the first step, but not the second, which is the application of observational analysis statistics. So far our data studies are best characterized as data screening; in fact, a very limited type data screening. Such screening efforts must be expanded to find all the good questions, and to find them before they become political issues. We also must recognize that data screening only raises the question. A new effort in observational analysis, unique to EPA in general, is necessary to find valid answers from existing data.

Because of the massive archival data at our disposal, there is another important principle that should not be overlooked. This principle, once the archival data raises the question, is to collect new data specifically to answer that question or hypothesis. It is, of course, easier to analyze existing data than to design and execute a valid experiment, but the effort may be necessary to obtain the truth. In general, it is a good research practice to collect specific new data. EPA needs to emphasize efficiency in finding potential problems, and to initiate a new awareness of the need for reaching valid, complete answers.

RAISING THE STATISTICAL ANALYSIS LEVEL OF ENVIRONMENTAL MONITORING DATA

By Wayne R. Ott

INTRODUCTION

The term *monitoring data*, as used in this paper, denotes routine measurements used to represent or describe the state of the environment. The definition includes measurements of environmental contaminants and other variables in air and drinking water as well as in lakes, rivers, and marine waters. Routine monitoring data are collected at great expense and in large quantities by environmental control agencies.

An example of monitoring data is data generated by a metropolitan air-monitoring network. A general urban monitoring network may consist, for example, of 10 stations, each measuring the six "criteria" air pollutants: sulfur dioxide, nitrogen dioxide, ozone, hydrocarbons, carbon monoxide, and total suspended particulate. Five of these six pollutants usually are measured continuously on an hourly basis, and one is measured on a 24-hour basis. Potentially this gives 438,000 values per year for the hourly data (8,760 hours/ year x 5 pollutants x 10 stations) and 3,650 values per year (365 days x 1 pollutant x 10 stations) for the 24-hour data. The cost of installing and maintaining such a network can be substantial, over $60,000 initial investment per station and probably more than $100,000 per year for overall maintenance for all stations. Therefore, such a network has an original cost of more than $600,000 and can generate 441,650 values per year, about four values per dollar of annual operating cost. The metropolitan air monitoring networks installed across the Nation represent a national investment of possibly over $30 million.

These data ultimately find their way into large monitoring data archives, such as STORET, SAROAD, and the EPA water supply data bank. It is reasonable to ask whether the resources expended on the analysis, interpretation, and display of these data are sufficient (and whether the quality of these analyses is adequate) relative to the resources originally spent to collect the data.

THE PROBLEM

For the most part, the problem is that State and local air and water pollution control agencies do relatively little in-depth analysis of these data. Some agencies lack access to computers, and some even lack the technical expertise to perform in-depth statistical analyses. The usual practice is to calculate means and standard deviations, to note how many values are above or below environmental standards, and to report the data to the public in some form of environmental index. Very few agencies examine correlations between environmental variables and causative factors, for example, or study trends using formal techniques such as time series analysis, or probe underlying statistical characteristics of the data. In addition, EPA's effort to carry out in-depth monitoring data analysis is very limited in scope.

In one sense, we have the spectre of a vast resource, environmental monitoring data, that is largely unexploited and underanalyzed. Looking at the total picture, there appears to be undue emphasis on making measurements and minimal emphasis on interpreting the measurements once they are made. The goal appears to be to collect, then to store, but not really to analyze.

AREAS REQUIRING EMPHASIS

There are at least three areas where we need to strengthen our efforts in upgrading statistical and mathematical levels: examination of underlying statistical properties of the data, correlation analyses, and trend analyses.

Underlying Statistical Properties

Except for very simple curve fitting, no extensive work has been undertaken to examine the underlying distributions from which environmental quality data arise, either for air or for water quality data. This lack of detailed analysis is of particular concern because some regulatory decisions and environmental standards are based on the assumption that measured concentrations have a particular distribution, such as the lognormal distribution. In a partial effort to fill this need, the Office of Monitoring and Technical Support has completed contractual work to develop univariate curve-fitting software. For example, the computer program MODEL.FIT, still under testing, is a general-purpose tool to evaluate the suitability of different probability models, such as Gamma, lognormal, Weibull, and normal, for a given data set. More work must be done in applying such tools to existing monitoring data.

Correlation Analysis

At present there also appears to be too little emphasis on examing the correlations of important causative factors. In analysis of air, for example, we need to examine the correlations of meteorological variables, such as fuel use patterns and industrial and vehicular emissions, with measured pollutant concentrations. Some examples of recent questions of interest are:

- How did measured carbon monoxide and sulfur dioxide concentrations change in relation to last year's fuel shortage?
- Is the decline of sulfur dioxide concentrations in New York City over the last 10 years consistent with the predictions of diffusion models or proportional rollback models, which often have been criticized and on which major regulatory decisions are based?
- Can these complex phenomena be treated by multiple regression analysis, or are other mathematical tools more appropriate?

Such studies, if carried out in depth, may give EPA policymakers and managers important insights into the way regulatory and energy policies are affecting environmental quality.

Trend Studies

Some trend reports are now routinely produced by EPA, and they are excellent contributions to the environmental literature. However, these reports do not probe very deeply into trend phenomena of the data. For example, the Box-Jenkins[1] time series analysis is a powerful, but little used, tool for detecting underlying changes over time that may be masked by meteorological and random fluctuations. Stratification of observed values, by different meteorological or hydrological conditions, is another neglected method of analysis. Autocorrelation functions for existing data sets, which can reveal subtle details about concentration variation over time and are important for short-term forecasting, are not routinely calculated by EPA or State and local agencies.

CONCLUSION

Water quality data actually may have suffered from a somewhat greater lack of in-depth analysis than air quality data, not only because there are many more environmental variables in water than in air but also because there are less continuous data available. In both air and water analyses, however, the need exists for improved data interpretive techniques and for greater use of these techniques to detect subtle trends, to improve the display of data, and to translate the data into forms that are understandable to managers and policymakers.

In summary, a critical problem appears to exist in the field of environmental monitoring, a gap between the effort expended to collect the data and the effort expended to analyze and display it. As the EPA quality assurance effort moves increasingly toward the day when monitoring data will be of uniformly high quality, we may find ourselves in the embarrassing position of being unable to confidently determine whether the environment has grown better or worse, or the responsible factors, simply because we have not sufficiently analyzed the monitoring data. Certainly our efforts to improve the quality of the data must not outstrip our efforts to intelligently analyze it.

Finally, there are, I feel, two important needs in the area of monitoring data analysis: (1) a need for the Agency to produce and demonstrate more tools to aid in the analysis, interpretation, and display of environmental data (guideline documents, customized computer programs, statistical manuals), and (2) a need for the Agency to give greater emphasis to applying these tools and to carrying out high quality statistical and mathematical analyses of monitoring data.

REFERENCES

1 Box, George E.P. and Jenkins, Gwilym M., *Time Series Analysis, Forecasting, and Control, Holden-Day, Inc., San Francisco, 1970*

QUALITY ASSURANCE FOR ADP
(AND SCIENTIFIC INTERACTION WITH ADP)

By R.C. Rhodes

Quality assurance personnel and statisticians are also concerned about ADP. Quality assurance practitioners (and quality control statisticians) are interested in the producing end of the total process—obtaining good data for storage in the computer, while applied statisticians and scientists are concerned with the using end of the total process—performing good analyses on the data in the data bank.

QUALITY ASSURANCE

A total quality assurance system for pollution monitoring organizations involves the following elements presented in Table 1.

Table 1
Elements of a Quality Assurance System

- Quality policy
- Quality objectives
- Quality organization and responsibility
- QA manual
- QA plans
- Training
- Procurement control
    - Ordering
    - Receiving
    - Feedback and corrective action
- Calibration
    - Standards
    - Procedures
- Internal QC checks
- Operations
    - Sampling
    - Sample handling
    - Analysis
- Data
    - Transmission
    - Computation
    - Recording
    - Validation
- Preventative maintenance
- Reliability records and analysis
- Document control
- Configuration control
- Audits
    - Onsite system
    - Performance
- Corrective action
- Statistical analysis
- Quality reporting
- Quality investigation
- Interlab testing
- Quality costs

Some of these elements may not be as important for research programs as for monitoring efforts. In either endeavor, however, the product of pollution measurement programs is *data*. The quality of the data may be measured by:

- Accuracy
- Precision
- Completeness.

In the simplest definitions, accuracy is the closeness of the measured values to the truth, precision is the measure of repeatability, and completeness is a measure of the amount of data obtained.

A unique feature of data as a product is that examination of the product itself does not reveal its quality with respect to accuracy and precision. Completeness can be indicated only by the amount of data obtained compared with the amount of data which should have been obtained under perfect conditions.

Accuracy and precision must be determined from ancillary information obtained during the measurement process. Accuracy may be checked by using primary standards (standard reference materials from the National Bureau of Standards) or secondary standards traceable to these primary standards, for calibration. Other information relative to accuracy may result from interlaboratory comparisons.

Checks for precision are obtained during the measurement process for internal quality control by using duplicate samples, spiked samples, interspersed standards, and so forth. The use of back-to-back duplicates of split samples in the analytical portion of a measurement system provides some internal quality control, although it is an inadequate (ultraconservative) measure of precision for the entire measurement process. One of the best methods of measuring the precision of the overall measurement system is the use of dual or colocated sampling with the analyses of the two samples being made as independently as possible. Those responsible for measurement systems are strongly encouraged to use, at least to a limited extent, colocated sampling to obtain overall system precision estimates.

Apparently, none of the pollution data banks currently incorporates, either directly or indirectly, measurements of precision and accuracy which can be matched with given blocks of data in the data banks. (Some efforts are being made in this direction.) Certainly, a key question in any enforcement situation is: "What objective evidence (data) exists which assures or measures the accuracy and precision of the data on which the enforcement action is based?"

Completeness of data sets is important in research efforts where the objective is to relate pollution data to other data, such as health effects or meteorological information. For example, complete health effects data are of little value if no matching air pollution data exist. Obviously, completeness of data is important in compliance monitoring.

Ideally, precision and accuracy data should be available (i.e., reported) together with the monitoring data of the data banks so that confidence limits may be applied to each of the monitoring data values. Another type of auxiliary data which should be included in the data in the data banks is a "special events file." Unusual pollution measurements may result from special events or circumstances observed or recorded at the time of sampling. Such information would be helpful to users of the data banks.

The use of ADP systems and quality assurance systems should be mutually beneficial. Obviously, ADP can and should be used for statistical computations, monitoring data summaries, and pertinent record-keeping aspects of quality assurance. On the other hand, quality assurance concepts and techniques can and should be used in procuring and operating ADP systems. Particularly applicable to integrated analysis/computer facilities and operations are the quality assurance elements presented in Table 2 which have been selected from the list in Table 1.

All of the quality assurance elements presented in Table 2 except perhaps for "Calibration and internal QC checks" apply equally well to ADP data bank/central computer procurement and operations. (Even for these systems, the periodic use of test programs may be considered a form of calibration and internal quality control.)

Table 2
Quality Assurance Elements Particularly Applicable to ADP

QA Elements	Specific Activities and Considerations
Procurement	
Ordering	Specifications
Receiving	Performance demonstration Operating and maintenance manuals Operational computer programs Warranty
Receiving	Complete operational checkout
Calibration and internal QC checks	Acceptability criteria and corrective action procedure Filing of results and traceability to monitoring data
Reliability records and analysis	Recording of frequency and cause of failure, MTBF for system, and components
Preventive maintenance	Establishing optimum schedules and recording evidence of maintenance actions
Document control Operational procedures Computer programs Configuration control	Are records adequate for procedures, software, and hardware configuration to be reconstructed for some specific past date?
Configuration control	
Data validation	
Manual editing	Techniques of detecting human errors
Scientific validation	Techniques involving scientific considerations to detect questionable data Spatial and temporal continuity Interrelationships between different measurements

STATISTICAL ANALYSIS OF DATA IN DATA BANKS

Potential users whose work should be greatly enhanced by data banks and ADP capability ask the following questions:

- What data are in the data banks?
  - Parameters
  - Time
  - Geographic location
- What data analysis programs are operational?
  - Statistical
  - Mathematical
- How to *talk* to the big computers without learning a new language such as FORTRAN?
- What interactive graphics programs are operational for use with a CRT, so that the data can be seen in various ways before, during, and after analysis?
  - Histograms without transformations
  - Histograms with transformations
  - Correlation (X-Y) plots

- X-Y-Z plots with different symbols to indicate levels of Z
- Time (chronological) plots, one or more variables on same plot
- Plots of residuals
    - Regression
    - Analysis of variance
- Distribution of differences and percent differences for paired data
- Map arrays (given coordinates and data values)
- Contour computations, plots, and map overlays
- Provision for deletion of specified data points prior to computations

. How and when will the capability exist to do all these things?

What non-ADP scientific users need to get optimum use of ADP are:

INstructions for
Scientific and
Technical
Users
(IN SITU)

or

Simple
Understandable
Instructions for
Technical
Scientific
Users
(SUITSU)

in English (not computerese or FORTRAN). All conversation between the user and the computer should be in English or mnemonic English abbreviations.

SUMMARY

Two major concerns related to automated data processing (ADP) systems are quality assurance of the data which enter the data system and scientific interactive use of data after they have been stored in the system.

In the past, much of the data entering the pollution monitoring data banks was generated through non-ADP systems; i.e., manual analytical methods and data gathering and reporting methods. Furthermore, quality assurance generally has dealt with the complete measurement-system-sampling, analysis, and data validation (Figure 1), rather than with ADP for the data bank.

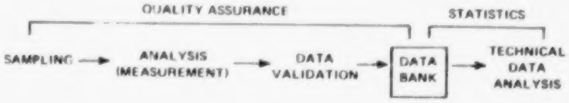


Figure 1
Manual Measurement System with ADP Data Bank

Recently, an increasing number of analytical pollution measurement methods have been automated with micro- and minicomputers. These also may be used as temporary data banks and may be directly tied in with the larger computers handling the master data bank (Figure 2). This makes it necessary for the quality assurance process to focus more on the ADP for the data producing system, rather than on the data storing system alone.

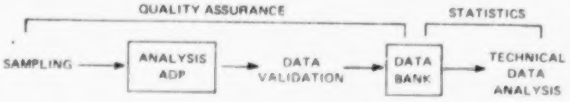


Figure 2
Integrated Analysis-ADP System with ADP Data Bank

Potential users of the data banks, such as statisticians and other scientists who could benefit from using the data and who desire to make the best analysis of the data stored in the bank, are reluctant to do so because of the complicated learning process.

HOW TO WRITE BETTER COMPUTER PROGRAMS

By Andrea T. Kelsey

Poor computer programing is a weakness found in statistical data analysis, as well as in all ADP environments. In general, computer programs are poorly designed, poorly coded, and hardly ever bugfree. In addition, they are usually impossible to understand and nearly impossible to maintain and to modify.

Good programers are hard to find, and even when good, they are often at a loss to teach their successful methods to others. Data analysts are still awaiting the packages or the language that will enable them to analyze data properly, and with a reasonable amount of effort. For them to have this software, and for it to be correct and maintainable, the quality of computer programing must be improved.

Structured programing appeared to be the answer to the problem of poor computer programing. The two basic rules of structured programing are: (1) limit the choice of constructs which the programer can use, and (2) break the program down into modules which are functionally independent.

Although they are sensible, these rules tell nothing about how to write a structured program, only what one should look like. Specifically, structured programing does not answer the question, "Which collection of modules is right for this particular problem?"; that is, "What should the program structure be?"

In the book *Principles of Program Design*, Michael Jackson[1] describes a programing technique which shall be called MJT for the purposes of this paper. This technique does answer the question, "What should the program structure be?" MJT is most applicable for programs concerned with data processing and with the development of systems. A scientific data analysis system can readily be described as a data processing program with sophisticated elementary operations. Therefore, MJT would be especially valuable in the development of any comprehensive data analysis system.

Basically, MJT consists of a formal analysis of the structure of the data files on which the program must operate. A program structure is then designed which has the same "shape" as the data. The machine-executable instructions are allocated to the program structure where necessary to accomplish the objectives of the program. All the decisions are made before any coding is done.

There are five steps to writing a program using MJT. They are described below:

1. Draw a data structure for each input and output file. Use any combination of the three types of diagrams (sequence, selection, and iteration) shown in Figure 1. Figure 2 shows examples of two data structures. The input file contains personnel records sorted by department, with each record containing department code, name, salary, and status. The status contains either a 'C' for current employee or an 'F' for former employee. The second data structure is the output report. It contains a report heading followed by a report body. The report body contains an iteration of lines. Each line contains the name, followed by the salary, followed by the status. The status is either 'CURRENT' or 'FORMER' depending on the value of the input status.

2. Draw the program structure. In this step, look at the data structures and draw all one-to-one correspondences between the input and output files. As a result, the program structure can be drawn. The one-to-one correspondences in the example are:

   - For each file there is one report, one heading, and one report body.
   - For each input record there is one line on the report.
   - For each name, salary, and status on input there is a name, a salary, and a status on output.
   - For each 'C' for status on input there is 'CURRENT' on output.
   - For each 'F' for status on input there is 'FORMER' on output.

The resultant program structure is found in Figure 3.

3. List and allocate elementary operations. In this step list all machine-executable instructions necessary to accomplish the program objective and allocate these operations to the program structure. In the example, the list contains:

1 Read record
2 Write heading
3 Write line of report
4 Open file
5 Close file
6 Stop run
7 Move NAME to output line
8 Move SALARY to output line
9 Move 'CURRENT' to output line
10 Move 'FORMER' to output line.

The allocation of the elementary operations is shown in Figure 4.

4. Write the schematic logic. Take the program structure and write it in a special language that allows only the three constructs shown in Figure 1:

- Sequence
- Selection
- Iteration.

The schematic logic for the example is:

```
PFILE seq
        Open FILE; read FILE;
    PHEADING seq
            Write heading;
    PHEADING end
    PBODY seq
        PRECLINE iter until end of file;
            PNAME seq
                move NAME to line;
            PNAME end
            PSALARY seq
                move SALARY to line;
            PSALARY end
            PSTATUS select (CURRENT)
                move 'CURRENT' to line;
            PSTATUS or (FORMER)
                move 'FORMER' to line;
            PSTATUS end
            Write line;
            Read FILE;
        PRECLINE END
    PBODY end
        Close FILE;
        Stop run;
PFILE end
```

5. Write the code. Take the schematic logic and write the code in any programing language.

The result of MJT is a correct program, one that has the right structure so that modifications can be made without causing the unwanted interactions that usually accompany a modification. The example is a simple one, but the advantage of MJT is that when it is used on a large, complicated program, the program becomes a series of simple processes.

Other aspects of MJT are briefly mentioned below:

- *Program Inversion*. There are times when the data structures will not fit together and when all necessary one-to-one correspondences cannot be found. This is called a structure clash. Sorting the files can sometimes resolve a structure clash. If not, then program inversion can be a solution. The technique involves the following two steps:
  - Think of the program as two programs. The first program has an output file which resolves the structure clash. This output file is an input file to the second program.
  - Convert one program so that it can run as a subroutine of the other. This is easy, once the schematic logic has been written for both programs. A cookbook method is used to convert one program to a subroutine.

- *Backtracking*. In some cases where a selection is necessary, the condition of the selection cannot be determined at the selection time. Backtracking means assuming one of these selections is correct, and using it. If it is found to be incorrect, one branches to the correct selection.

- *Optimization Rules.* There are two rules on optimization:
  - Don't do it.
  - For experts only: Don't do it until you have a perfectly clear and unoptimized solution.

  The main reason is that optimization makes a system less reliable, harder to maintain, and therefore more expensive to build and operate.

- *GO TO Statement Use*. The conventional objection to using GO TO statements is that they permit unrestrained branching from one part of a program to another. MJT does not allow this use of the GO TO, either. But in three cases, MJT allows GO TO statements:
    - In backtracking to branch to the correct selection.
    - When doing program inversion (restricted use of a computed GO TO).
    - When the shortcomings of the programing language force its use to implement the schematic logic.

  Read Ahead Principle. MJT always uses this principle. It is defined below:

  Always have one read operation immediately following the open operation for each input file. Then always read another record when the previous record has been completely processed. By following this principle, the logic of when to read does not become a problem.

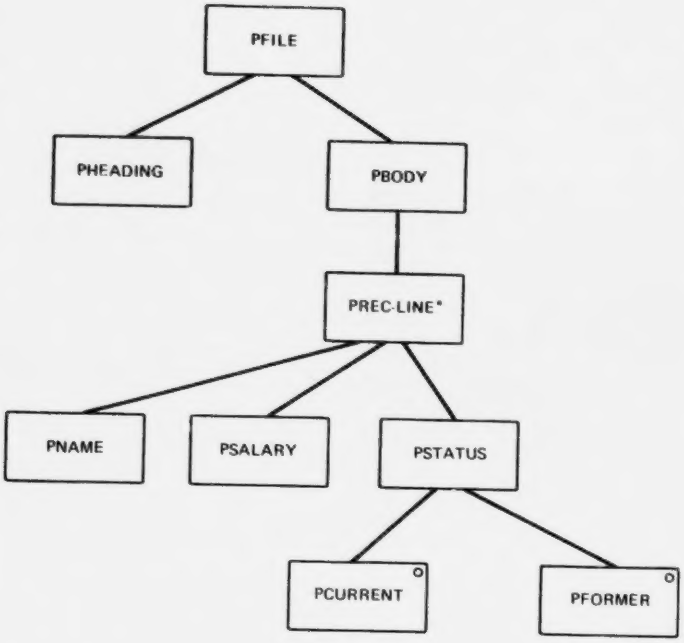
- *Collating*. Before learning MJT, matching two or more sorted input files has almost always been a problem that must be solved each time a collating program is needed. MJT teaches that there is only one collating problem, which always has the same solution. This solution, made possible by the read ahead principle, requires too detailed a description for this paper. It is mentioned only to point out that MJT has a mechanical solution for collating problems once the collating keys have been identified.

My experience with this technique has involved a large data processing program which was written in COBOL and which interacts with a System 2000 data base. By using MJT, the programer understands the program well and is convinced that the program is correct. When modifications are needed, the programer knows exactly where to make the changes. The programer knows, also, that the changes will not adversely affect the other parts of the program.

In conclusion, the use of this technique would greatly enhance the quality of the software the data analyst has to work with, and would make development of a comprehensive data analysis package reasonable.

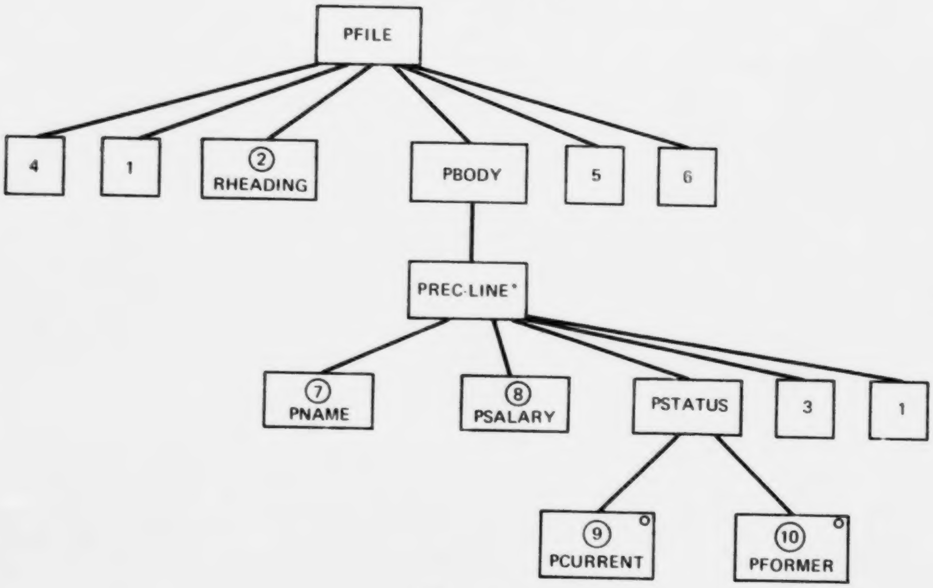
REFERENCE

1 M. A. Jackson, *Principles of Program Design*, Academic Press, 1975.


PROGRAM STRUCTURE

PFILE
PHEADING
PBODY
PREC-LINE*
PNAME
PSALARY
PSTATUS
PCURRENT o
PFORMER o

Figure 1


PROGRAM STRUCTURE

PFILE
4
1
② RHEADING
PBODY
5
6
PREC-LINE*
⑦ PNAME
⑧ PSALARY
PSTATUS
3
1
⑨ PCURRENT o
⑩ PFORMER o

Figure 2

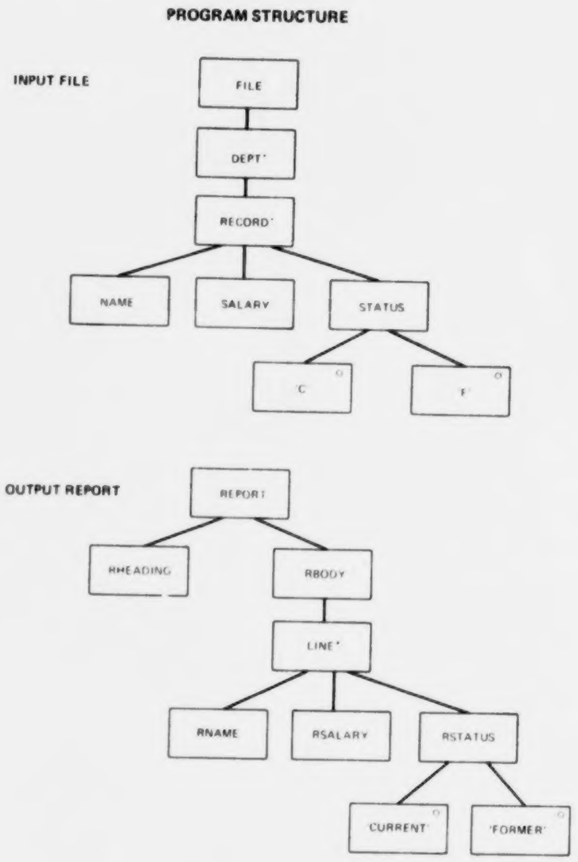

PROGRAM STRUCTURE
INPUT FILE
FILE
DEPT*
RECORD*
NAME
SALARY
STATUS
'C' o
'F' o
OUTPUT REPORT
REPORT
RHEADING
RBODY
LINE*
RNAME
RSALARY
RSTATUS
'CURRENT' o
'FORMER' o

Figure 3

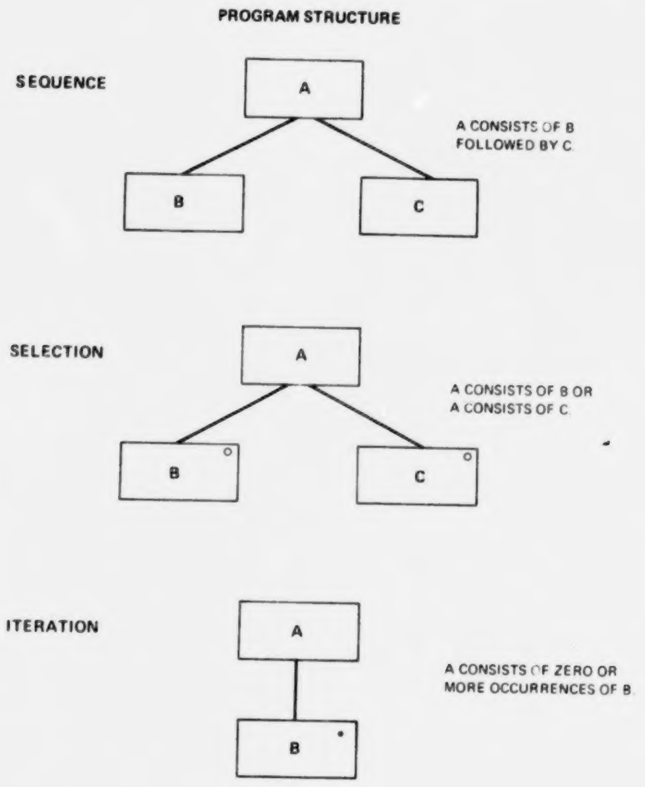

PROGRAM STRUCTURE
SEQUENCE
A
B
C
A CONSISTS OF B
FOLLOWED BY C.
SELECTION
A
B o
C o
A CONSISTS OF B OR
A CONSISTS OF C.
ITERATION
A
B *
A CONSISTS OF ZERO OR
MORE OCCURRENCES OF B.

Figure 4

SUMMARY OF DISCUSSION PERIOD - PANEL III

In the discussion following the presentation concerned with strengths and weaknesses of analysis of scientific data, the following conclusions were drawn.

Data System Uses

The panel agreed that EPA data systems accomplish more than just storage of data. Research data, such as health effects data, are being subjected to extensive analysis, and resources are being provided specifically for that purpose. In the case of most routine compliance monitoring data, so little or the right kind of analysis is being done that, in effect, only data storage is accomplished. There was a rejoinder from the floor that perhaps it should not be the function of the monitoring data systems to provide the analysis but merely to make the data available and that other components of EPA should be concerned with the analysis. There was sympathy for this view among some of the panel. The panel generally agreed that these systems should provide for easy retrieval of the data in a form which can be analyzed readily.

Experimental Design and Analysis

Some felt R&D management is not insisting that scientists apply good experimental design techniques. Various members of the panel felt that the kind of people required to do the statistical design and subsequent analysis of experiments generally is not available within ADP. (Notable exceptions are the HERL and EMSL at RTP which have their own in-house cadre of statisticians.) Such resources, therefore, must be provided either from within the R&D laboratories or by contract. In the case of contracts the panel felt that an employer-employee relationship was needed in this situation. How one stimulates such a relationship, given the contracting process was asked. Several panel members said "not very well." Others said some sort of regional contract is needed which would allow the contractors to become so familiar with the research program that they could contribute to the design and analysis of experiments.

Quality Assurance Guidelines

In answer to what ORD is doing to improve monitoring QA it was pointed out that over the past year or so, there have been a series of "Guidelines for QA Programs" for various pollutant measurement methods (ambient and source air). There are a dozen or so currently available and more are being developed. What is needed most is a general quality assurance handbook, "QA Handbook for Air Pollution Measurement Systems," which will be ready, hopefully, by February 1976. While written specifically for air programs, much of it will be applicable to any media. It was noted from the floor that there has been a quality control handbook for water available for 4 years. Either people ignore it or, after using it, data is put in the data banks without distinguishing whether good quality control was used in the production of that data or not. The panel concluded that the water handbook was important and good, but that it was directed primarily to the analytical laboratory, while the real need is to address the total measurement process.

Appreciable resources are being devoted to some aspects of the quality assurance effort, but applications of quality control techniques to data once it is in the data bank is recent and quite limited. There are substantial policy questions involved with ORD's role in this area. ORD is a technical organization and can provide such things as software, studies, and techniques, but STORET people for example, would be the ones applying such techniques to STORET data.

Quality Assurance Responsibilities

The panel agreed that EPA's quality assurance efforts should not be transferred to the program offices for various technical reasons, such as coordination with State agencies which cannot afford to be split this way. Any program in which data are being generated should implement quality control methods and techniques and should have persons assigned to quality assurance responsibilities. However, because of the many common quality assurance activities across media and programs, the top level quality assurance organization for EPA should be in one place for maximum efficiency and effectiveness.

QA Information in Data Bases

The panel agreed that the quality assurance information that should be carried in the data base should at least include a precise description of the conditions under which the data were derived. Also, some measure of accuracy and precision associated with the data would be highly desirable. There was no agreement on whether explicit measures of these quantities should be carried in the data base or whether such measures should be related to the conditions under which the data were produced.

ADP Centralization

The general feeling among the panel was that it would be good to centralize in one office or laboratory all ADP people at a particular facility. As with many other resources, some combination of centralization and decentralization is appropriate. A laboratory or office which has substantial ADP needs, such as RTP, should have its own ADP staff.

Policy for Non-Agency Requests

The panel stated that the Freedom of Information Act is the EPA policy toward providing data to people outside the agency.

Contract for Statistical Data

The consensus of the panel was that while the worth of a statistical services contract could be debated at length, any such contract should be negotiated by the research organization itself, not MIDSD.

Structured Programing

Should GO TO statements be used and if so when?

The problem with the GO TO statement was expressed as permitting unrestricted movement among components of the program. However, limited use of the GO TO is called for in a few situations; namely, those identified in the paper on the Michael Jackson techniques.

It was decided that existing systems should probably be rewritten using structured techniques whenever modifications are extensive at all.

The Michael Jackson technique was said not to be a competitor of Structured Programing. The MJT answers the crucial question of how to construct a particular program. The result of its answer is a structured program.

Statistical Training

It was asked whether any efforts are being made to train chemists and other non-ADP people in how to use the statistical packages. Several seminars were described which have been open to everyone with a need to know, but no concerted effort specifically directed towards non-ADP people was known.

Scientific Software

One way for the RTP-NCC staff to get information on software and training needs was established at Research Triangle Park before the National Computer Center began. It is the Scientific Software Committee. Whether it should be expanded to include the regions or other groups is being discussed. Some sort of formal means of polling the user community on scientific software needs is definitely needed.

To some degree, the panel advocates providing users with custom software packages to meet their needs. In developing a software package, a class of users should be identified and the package should be tailor-made for that class. But the classes should be kept very large.

Statistical Package Development

Whether the development of a comprehensive statistical package is something that must be done in-house or can be contracted was discussed. The definition of need must be done in-house up through having functional and performance specifications. Only then can the development of the package be contracted.

It was then asked whether the development of comprehensive statistical packages should be the responsibility of a central group or of the data base managers.

The panel stated the data base managers should determine needs; the central group should be the means of seeing that the needs are met.

THE UTILITY OF BIBLIOGRAPHIC INFORMATION RETRIEVAL SYSTEMS

By Johnny E. Knight

Over the last several decades we have seen what is now popularly referred to as an "information explosion." Presently, the United States spends at least $11.8 billion a year on scientific and technical information activities. As much as $6.1 billion of this amount was spent on all distribution-associated aspects including distribution, storage, and retrieval. In the past decade, the number of scientific periodicals has increased by 9 percent while technical report literature has increased by an estimated 16 percent.[1]

At least four facts have made it almost inevitable that some form of computer bibliographic data handling system would be developed. First, the amount of scientific and technical literature has become massive. Second, the assimilation of information by scientists and engineers, primarily as salary costs attributable to browsing, searching for information, and reading is reported to be as much as $3.3 billion.[1] A cost reduction in this respect would be greatly desired. Third, new computer hardware and software technology have reduced literature search costs significantly. For example, in 1965 the search cost of a 500,000 record file was approximately $1,000. Today, the search cost for the same file is approximately $10.[2] Fourth, the present day requester of information has an almost fanatical desire to know answers instantly.

In their book on online information retrieval systems, Lancaster and Fayen[3] have summarized hardware developments over the years as follows:

- Before 1940: mostly card catalogs and printed book indexes.
- During 1940-1949: the first application of semimechanized approaches, including edge-notched cards and the optical coincidence (peek-a-boo) principle; the microfilm searching system (the Rapid Selector).
- During 1950-1959: the first fairly widespread use of punched card data processing equipment; some early computer systems; further microimage searching systems.
- During 1960-1969: more general application of digital computers to information retrieval in an off-line, batch-process mode; some experiments with online, interactive systems; more advanced microimage searching systems.
- From 1970 to the present: definite trend toward design of online systems and conversion of batch systems to the online mode.

As the summary shows, until recently all bibliographies had to be compiled and edited by hand primarily from printed indices. This laborious task was extremely time-consuming and precluded the "immediate" response time generally expected today by our "modern researchers" fighting those so-called "hot" issues, which have a tendency to appear overnight.

One of the first attempts to use computers for bibliographic literature search preparation was in 1966 by the National Library of Medicine (NLM) with the Medical Literature Analysis and Retrieval System (MEDLARS). This was a batch-oriented system requiring the user to send his request to NLM. Delay in receiving a reply was sometimes considerable.

It was soon recognized that the batch mode of searching also had other deficiencies. Generally, the requester had to rely on an information specialist to conduct his search. The search could not be developed heuristically, and the user lost another very important option: browsability. Even manual methods allowed these two aspects of searching.

Of course, the answer is immediately obvious to those in the ADP field. Online real-time search systems would eliminate these inadequacies. In 1965, there were about 20 machine-readable data bases. Today there are approximately 200 data bases available to the public. About 50 of these are online systems. Thus, a trend begun in 1970 has become the established mode of operation in 1975.

Although computer storage and retrieval methods are beyond the scope of this paper, they should be mentioned briefly. Although other methods are available, either sequential or inverted file organization is most often used for bibliographic data systems. An organization method must be chosen to suit the computer hardware, the data characteristics, and use requirements of the data. Use of inverted files with indexed sequential access has the advantage of easy file maintenance and allows search strategy evaluation using Boolean logical connectors without actually retrieving the citation or document data. Recently, however, Gerald Salton has produced an alternative to the inverted file which he calls clustered file organization.[4] Salton is very critical of inverted files and states that his method requires less storage and allows more flexible searching than other methods.

Historically, large data bases have utilized as much of their particular computer facility as could or would be allowed. They were so expensive to develop, maintain, and make available that in some cases only Government funding allowed their existence. Lancaster and Fayen state that the computer specialist has been forced to extend computer technology to handle more and more data.[3] For whatever reason, it seems that industry journals announce each month new technology and innovations that allow more data in smaller spaces and faster retrievals.

At first, use of the data bases was free of charge; but as their use became more than a novelty, charging systems were implemented and Government subsidies were sought. Originally, the data bases were developed privately to be used in-house. When it became recognized that these systems had a wide user community, they were made available through commercial data centers such as the Lockheed Missiles and Space Company (LMS) and the System Development Corporation (SDC). Recently, third party vendors or brokers are beginning to make these systems available as self-supporting, profitmaking enterprises.

Although many search systems are commercially available, probably the two most generally available within the EPA, and possibly to the public in general, are the ORBIT system from SDC and the DIALOG system from LMS. Medical Literature Analysis and Retrieval System Online (MEDLINE) is also widely used within the Agency. Even though it is no longer maintained by SDC for NLM, its search system is essentially an ORBIT implementation.

Both DIALOG and ORBIT search languages accomplish the same end; that is, selection of a number of documents pertinent to a given topic based on a user-formulated search strategy. For demonstration purposes, suppose we wish to find citations/abstracts on the topic of: "How do sulfur oxides get into the Detroit River?" The DIALOG system was designed for indirect searching as illustrated on our hypothetical inquiry:

```
? S SULFUR; S OXIDES; S DETROIT; S RIVER?
      1  5324    SULFUR
      2  9760    OXIDES
      3   290    DETROIT
      4   851    RIVER?

? C 1 and 2 and 3 and 4

      5    53    1 and 2 and 3 and 4

? S SULFUR(1W)OXIDES); S DETROIT(W)RIVER

      6  3561    SULFUR(1W)OXIDES
      7    65    DETROIT(W)RIVER

? C 6 and 7

      8    28    6 and 7
```

The select (S) command causes numbered subsets of the inverted file to be set aside and displays the number of postings or "hits" for the selected key. The subsets are then available for use with other DIALOG commands. The combine (C) command allows Boolean logic to be used with the set numbers. The "?" is the computer prompt that it is expecting a command. Also, it may be used as shown in set 4 to indicate to the system that right-hand truncation of that key is desired.

Another feature of the system is shown in sets 6 and 7. This feature allows the user to request that the documents to be retrieved must contain the stated keys. Additionally, these keys must occur within a specified number of words of each other (string searching). This allows the user to request that the documents be more specific; that is, the key "sulfur" must occur within one word (disregarding insignificant words such as "the") of the key "oxides." Obviously, this would eliminate documents occurring in set 5 which might have discussed sulfur compounds and nitrogen oxides together with some other river besides the Detroit River.

The ORBIT system was designed to allow a searcher to enter his search strategy directly as illustrated:

```
SS1

      USER: SULFUR
      PROG: PSTG  (5324)

SS2

      USER: 1 AND OXIDES AND
            DETROIT AND RIVER
      PROG: PSTG (53)

SS3

      USER: STRS   :SULFUR  OXIDES:
      PROG: PSTG  (30)
```

The ORBIT system informs the user of the subset number it will next form (SS1). The system then prompts the user for his input by displaying "USER:". Its replies are always preceded by "PROG:" to indicate that it has control. Four separate sets may be selected and then combined into a fifth as shown with DIALOG. Alternatively, only one set with the final 53 postings need be created, whichever the user desires. This system also allows string searching as shown in SS3; however, a previous set of postings must be searched.

The indirect approach is more helpful to the inexperienced user, but the direct approach would probably be preferred by the more experienced user. Both of these commercial systems DIALOG and ORBIT, allow some form of display of the inverted file keys together with their statistics to help the user decide on his strategy as the search proceeds. A variety of terminal and off-line print formats are available. Both systems are used on "indexed" and "free-text" or "natural language" data bases, depending on certain fields declared when the data base is loaded.

Although it is impossible to state a monetary value for bibliographic data bases, they appear to be providing a useful service to researchers. Williams estimates that in 1965 there were only 10,000 users, whereas in 1975 there were over a million in the United States and Canada.[2]

Presently the EPA Research Triangle Park (RTP) library has approximately 30 online data bases available to be searched for its user community. On a monthly basis, the RTP library averages about 50 to 60 subject requests from about 175 individual researchers and spends approximately $2,000 for these inquiries. The usage of this library is an indicator of the value of bibliographic data bases to the general user community.

REFERENCES

1 Burchinal, L.G., "Recent Trends in Communication of TSI," *Bull. Amer. Soc. Information Sci.* 2(3):9, 1975

2 William, M.E., "Use of Machine-Readable Data Bases." Presented at 38th American Society for Information Science Annual Meeting, October 26-30, 1975, Boston, Massachusetts.

3 Lancaster, F.W. and Fayen, E.G., *Information Retrieval On-Line,* Los Angeles, California: Melville Publishing Co., 1973.

4 Salton, G., "Dynamic Document Processing," *Association for Computing Machinery Communications.* 15(7):658-668, July 1972.

BIOLOGICAL DATA HANDLING SYSTEM
(BIO-STORET)

By Cornelius I. Weber

The need for a functional, computerized system to handle data on the communities of indigenous aquatic organisms in the Federal water pollution control program has remained unfulfilled for nearly 20 years. The water quality data storage and retrieval system (STORET) developed in 1957 has been adequate for the storage and manipulation of physical and chemical data, but despite many improvements, it still lacks the ability to accommodate the hierarchical structure of biological data. The deficiencies in STORET have prevented the computerization and analysis of the bulk of the data from nearly 30,000 plankton samples collected during the operation of the National Water Quality Network, and from large numbers of other biological samples collected by current programs of the EPA and other Federal, State, and private agencies engaged in studies of the aquatic environment. Furthermore, many State programs have delayed the collection of biological samples until an adequate EPA computerized biological data handling system is available.

The mandate for the collection of data on communities of indigenous aquatic organisms contained first in Section 4(c) of the Water Pollution Control Act of 1956 (Public Law 660) was greatly expanded in the 1972 Amendments to the Federal Water Pollution Control Act (Public Law 92-500). This legislation contained direct or indirect reference to the need for the collection of biological data in at least 15 sections, and emphasized the need to restore and maintain the biological integrity of the Nation's waters, thus ensuring the protection and propagation of fish, shellfish, and wildlife. It also made numerous references to the need for the collection of data on the effects of pollutants on the diversity, productivity, and stability of communities of indigenous aquatic organisms, necessary to achieve the overall objectives of the Act.

In the spring of 1973, the staffs of the Monitoring and Data Support Division, Office of Water Programs, and the Aquatic Biology Methods Research Program, Office of Research and Development, initiated a project to develop a new computerized data handling system (BIO-STORET) for the storage, retrieval, and analysis of field and laboratory biological data. These data are concerned with the structure and function of communities of indigenous aquatic organisms and are currently being generated by the EPA as well as by other Federal and State agencies whose programs include inland, estuarine, and marine water quality monitoring, compliance monitoring, and studies of the effects of ocean-disposed wastes and heated water discharges. Such programs are mandated under Sections 104, 106, 308, 314, 316 and other applicable sections of the 1972 Amendments of the Federal Water Pollution Control Act. Communities of indigenous aquatic organisms provided for in the system include phytoplankton, zooplankton, meroplankton, periphyton, macroalgae, macrophyton, macroinvertebrates, and fish (Figure 1). It was agreed that the responsibility for the development of BIO-STORET rested with the Office of Research and Development and that, once the system became operational, it would be supported by the Office of Water Programs as a companion to the Water Quality File. Management of the BIO-STORET system was to reside in the Information Access and User Assistance Branch.

Contracts awarded in 1973 resulted in the preparation of a system requirements specification, a system design specification, and master taxonomic (6,000 species), parameter, and station files to be included in the initial system. An ad hoc steering committee, consisting of senior biologists representing a cross section of EPA regional and research programs and personnel from USGS and NOAA, was organized to assist the contractor in defining the system requirements and design. Further work on the system was delayed until 1975 because of the lack of funds.

The current contract with MRI Systems Corporation, Austin, Texas, was awarded in March 1975 for development of the detailed software design, software coding and debugging, and system implementation. BIO-STORET will utilize SYSTEM 2000, a generalized data base management software package marketed by MRI Systems Corporation, and available on OSI and the EPA Univac 1110 at RTP. The project schedule calls for the completion and implementation of the BIO-STORET system on OSI in the spring of 1976.

The preliminary design specifications for BIO-STORET in the new SYSTEM 2000 software environment were completed October 17, 1975 (Figure 2). Copies of the design specifications are available from Cornelius I. Weber, Aquatic Biology Section, Environmental Monitoring & Support Laboratory, U.S. Environmental Protection Agency, Cincinnati, Ohio 45268.

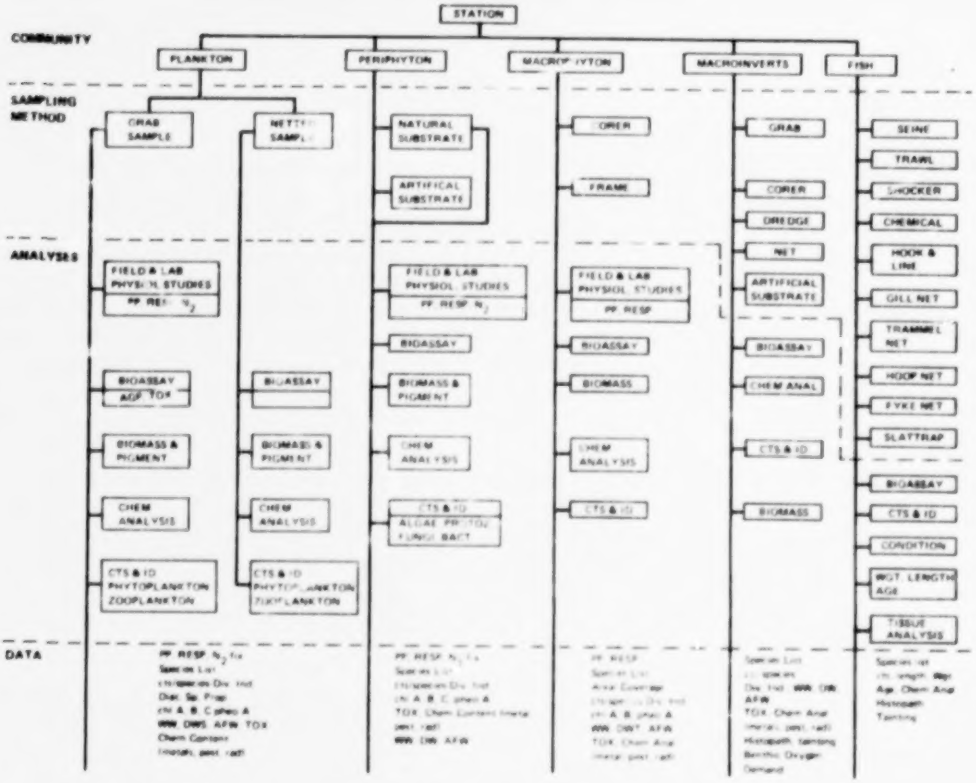

STATION
COMMUNITY
PLANKTON
PERIPHYTON
MACROPHYTON
MACROINVERTS
FISH
SAMPLING METHOD
GRAB SAMPLE
NETTED SAMPLE
NATURAL SUBSTRATE
ARTIFICAL SUBSTRATE
CORER
FRAME
GRAB
CORER
DREDGE
NET
ARTIFICIAL SUBSTRATE
SEINE
TRAWL
SHOCKER
CHEMICAL
HOOK & LINE
GILL NET
TRAMMEL NET
HOOP NET
FYKE NET
SLATTRAP
ANALYSES
FIELD & LAB PHYSIOL STUDIES
PP. RESP. N_2
BIOASSAY
AGP TOX
BIOMASS & PIGMENT
CHEM ANALYSIS
CTS & ID PHYTOPLANKTON ZOOPLANKTON
BIOASSAY
BIOMASS & PIGMENT
CHEM ANALYSIS
CTS & ID PHYTOPLANKTON ZOOPLANKTON
FIELD & LAB PHYSIOL. STUDIES
PP. RESP. N_2
BIOASSAY
BIOMASS & PIGMENT
CHEM ANALYSIS
CTS & ID ALGAE PROTOZ FUNGI BACT
FIELD & LAB PHYSIOL. STUDIES
PP. RESP
BIOASSAY
BIOMASS
CHEM ANALYSIS
CTS & ID
BIOASSAY
CHEM ANAL
CTS & ID
BIOMASS
BIOASSAY
CTS & ID
CONDITION
WGT. LENGTH AGE
TISSUE ANALYSIS
DATA
PP. RESP. N_2 fix
Species List
cts/species Div. Ind.
Diat. Sp. Prop.
chl A, B, C pheo A
WW, DWS, AFW, TOX
Chem Content
(metals, pest, rad)
PP. RESP. N_2 fix
Species List
cts/species Div. Ind.
chl A, B, C pheo A
TOX, Chem Content (metal,
pest, rad)
WW, DW, AFW
PP. RESP
Species List
Area Coverage
cts/species Div. Ind.
chl A, B pheo A
WW, DWT, AFW
TOX, Chem Anal
(metal, pest, rad)
Species List
cts/species
Div. Ind., WW, DW,
AFW
TOX, Chem Anal
(metals, pest, rad)
Histopath, tainting
Benthic Oxygen
Demand
Species list
cts, length, wgt
Age, Chem Anal
Histopath
Tainting

Figure 1
Types of Biological Data Accommodated by BIO-STORET

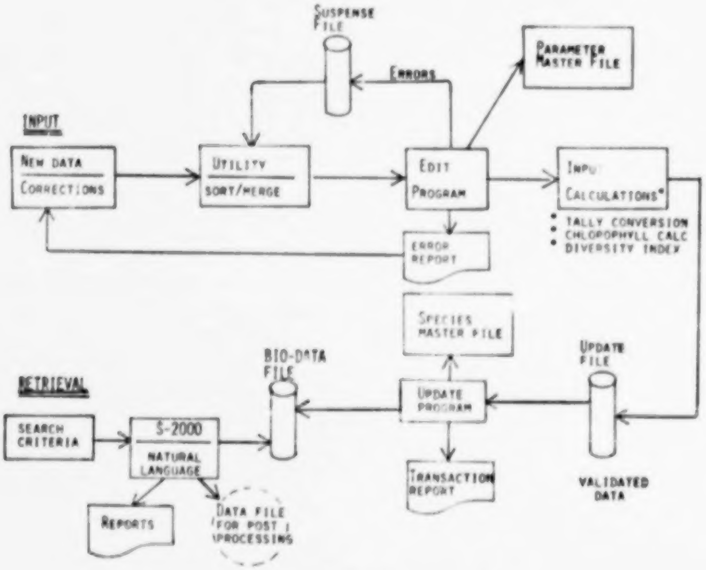

SUSPENSE FILE
ERRORS
PARAMETER MASTER FILE
INPUT
NEW DATA
CORRECTIONS
UTILITY
SORT/MERGE
EDIT PROGRAM
INPUT CALCULATIONS*
* TALLY CONVERSION
* CHLOPOPHYLL CALC
* DIVERSITY INDEX
ERROR REPORT
SPECIES MASTER FILE
UPDATE FILE
BIO-DATA FILE
RETRIEVAL
SEARCH CRITERIA
S-2000
NATURAL LANGUAGE
UPDATE PROGRAM
TRANSACTION REPORT
VALIDATED DATA
REPORTS
DATA FILE FOR POST PROCESSING

Figure 2
BIO-STORET System Flow Chart

UTILITY OF STORET

By C.S. Conger

Webster's Dictionary lists several definitions of the work "Utility," including: "being of a usable but inferior grade," "capable of serving as a substitute in various roles or positions," "kept for the production of a useful product rather than for show or as pets," "designed for general use," "the quality or state of being useful," "something useful or designed for use." STORET fits all the above descriptions and even a few others, at different times. However, it would be best if all but the first applied to STORET.

"Who, what for, and how" are questions which can be asked about STORET. There is no way to give a good representation of "how" in a short time, so this discussion will describe the "what for" and "who" of STORET.

The "what for" might be broken into basic functional areas of interest, which include: data management, water quality assessment, water quality management, and water pollution control program management.

Management and operation of STORET is primarily a data management function that should support the water program. Responsibilities for data management include providing efficient and reliable software that meets user requirements with regard to types of data accommodated, adequate controls for data entry and data quality, and appropriate analytical and retrieval routines. User training in the system is included in this responsibility. Data management adapts to the programs and activities it functions to support. It is not a "stand alone" activity with a mission of its own; it is a support function.

The water quality assessment function is concerned with all aspects of collection and analysis of water quality samples and with reporting the findings and trends of these analyses. Areas of responsibility include monitoring strategies, method of sample collection, and method and quality control of analytical procedures. The monitoring strategies should represent the questions to be answered by the reporting function so that they can provide appropriate data. The water quality assessment function relies heavily on STORET as a data handling tool. However, the groups charged with water quality assessment are responsible for the data collected and the reports produced from the STORET data. As procedures and controls for data collection improve, the data in STORET will improve as well. In actual practice, the easy access to previously collected and analyzed data provides data for analysis of problem areas and for suggested improvement priorities. As the definition of questions improves, STORET retrieval capabilities will continue to be extended to answer these questions.

Water quality assessment plays a rather passive role since it does not directly have an impact on the state of the Nation's waters; however, water quality management plays an active role. Its functions include: the 303 Basin plans, which recommend approaches and priorities for pollution control (based on water quality assessment and abatement technology), the permit program, which establishes and monitors effluent limitations, and the standards activities, which provide guidelines for attainable water quality goals. Water pollution control actions are more direct than water quality management, which is an applied technology. As such, it is supported more by experience than by scientific procedures. Data handling techniques and data requirements for water quality assessment activities are much more precise and rigorous than those for water quality management activities. In many of the latter activities, the use of data replaces or enhances the intuition and experience that is basic to applied technology and engineering. The pooling of data and the utility mode of STORET operation strongly support the water quality management activities. The bulk of the responsibility for water quality management activities rests in the States, and the tasks are jointly performed by the Regions and the States. Both water quality assessment and water quality management would benefit from improved communication of needs and priorities among the various groups involved. Use of a common data base, such as STORET, could serve as a starting point for communication improvements.

Management of the water pollution control program is essentially the responsibility of those making decisions within the Agency. Water quality assessment activities support this function by measuring progress in improvement of the Nation's water quality, but these activities do not necessarily identify the actions or decisions which were most effective in producing this improvement. Identification may require a management information system that can correlate the type of actions and the expenditures in certain areas (specific basins, States, or regions) with the assessment of water quality for that specific area. Aggregation of data on the Agency level may not provide the detail needed for correlation of cause and effect. Nevertheless, the data handling requirements to support the program management activities would not be directly related to STORET. Knowledgeable analyses of the data within STORET would possibly provide insight into the water quality progress but STORET cannot answer questions on which programs have the greatest impact on pollution abatement and on where funding priorities should be.

Now let us combine more "what for" with "who." National-level users of STORET are primarily concerned with research and with water quality assessment. These users, and their respective functions, are included in the following groups:

- Office of Research and Development: ORD use involves all appropriate phases of research to understand the causes and effects of pollution, the majority of which is conducted under Title I of PL 92-500.
- Office of Water Planning and Standards: The principal use has been for preparing 305(b) reports submitted by the States, and for responding to queries from other agencies with a national perspective.
- Council for Environmental Quality: CEQ prepares annual reports on the environment.
- National Commission on Water Quality: NCWQ recently submitted a draft report on the technological aspects, as well as the social, economic, and environmental impacts, of achieving or not achieving the goals for 1983.

Regional and State roles overlap in many cases, particularly at the present stage of implementing PL 92-500. While activities related to water quality assessment are conducted on the regional and State level (305(b) reports and Surveillance and Analysis Divisions activities), primary emphasis is on water quality management activities such as:

- Establishing standards - Primarily a regional responsibility, but the States are required to make recommendations for changes.
- Permitting - Equally divided between the States and regions; approximately half of the States have permit authority. Review is a regional responsibility.
- Planning - Performed at several levels, starting with 303 Basin plans prepared by the States and reviewed by the regions. Various interstate commissions will be involved in areawide planning.

Significantly, STORET has been a major data handling tool supporting all these various functions on the national, regional, State, and local levels.

The division of responsibilities for the various components of STORET across organizational lines has had an impact on STORET system management effectiveness. The components, or functions, and the respective organizations responsible are summarized as follows:

- Operation and maintenance: Data Processing and User Assistance Branch, Monitoring and Data Support Division (MDSD), Office of Water and Hazardous Materials (OWHM)
- Policy decisions: Water Quality Analysis Branch, MDSD, OWHM
- Vendor for computer resources: Management Information and Data Systems Division (MIDSD), Office of Planning and Management (OPM)
- User (storage): monitoring and data collection activities are found in various divisions within the Regions, States, other Federal agencies, and Office of Research and Development (ORD). Monitoring Branch, MDSD, OWHM, is also responsible for other aspects of monitoring and data collection.
- User (retrieval): found in all divisions of the Regions, States, and ORD, but may not be the same group responsible for collection and storage of the data.
- Designated STORET contact: within the Regions, this individual may be located in the Planning and Management Division, the Water Programs Division, or the Surveillance and Analysis Division.

From this division of responsibilities, it is seen that decisions made in one group with one line of management have a major impact on several other groups. Problems of communication and management occur because of the fragmentation of functions, and this fragmentation detracts from effective utilization of STORET and from a consensus regarding its future priorities.

STORET plays a dual support role. "How" STORET does this is by providing utility support to the Regions' and States' water quality assessment and water quality management activities, and by providing a data base for national reporting of trends and progress. STORET serves a utility role both by sharing software (providing for efficient storage and retrieval of data) and by pooling data in a common format (providing access to data collected by others). The utility role is the more successful of the two functions. The other role, national reporting of trends and progress in water quality, is a complex and highly technical responsibility, particularly when performed at the national level without benefit of localized knowledge of data, such as seasonal variations, type and location of major sources of pollution, natural background, and other variables that can have a significant impact on analysis and interpretation of data. Improvement in the role of a national data base will depend to a large extent on a better definition of the questions to be answered.

Other major STORET uses are explained below.

- Over 75 percent of the retrievals from STORET are in direct support of PL 92-500. The three major functions supported are reporting (such as 305 (a and b)), planning (303 Basin plans), and surveillance (Section 104).
- Users (in the utility mode) retrieve data primarily from their own waterways as this coincides with their area of responsibility or jurisdiction.
- Usage of the data by the data generators (through direct access) contributes significantly to improved data quality. The generators then have a vested interest in maintaining quality in a specific data base since they will benefit from its use.
- Detailed reporting by the States for the 305(b) reports provides a detailed national inventory, when taken collectively.
- Pooling of data contributes significantly to the value of the data base for both utility use and national reporting use.
- The groups responsible for data collection and storage are not the only users of the data. Many groups are concerned with only retrieval and use of the data and may not have a voice in the priorities for data collection or monitoring.
- STORET is not a management information system. STORET retrieval requires technical analysis to produce a visible product or a base for management decisions.
- Increased usage appears to be related to increases in use of quantitative data for decisions previously made by intuition and experience.

From a review of the STORET mode of use, it is obvious that STORET has had a significant impact on improving water pollution control and abatement procedures. With good data handling, those national or local organizations concerned with water pollution control and abatement can choose to use scientific and technical information to make their decisions. Without good data handling capabilities, they have no choice. Taking a little liberty with Webster, another definition of utility would be: STORET Utility - Having the quality or state of being useful, designed for general use to serve various roles in water quality management, and kept for producing a useful product.

THE USES AND USERS OF AEROS

By James R. Hammerle

BACKGROUND

The United States Environmental Protection Agency's comprehensive air pollution information system, the Aerometric and Emissions Reporting System (AEROS), is a valuable tool for managing the national air pollution control program effectively. Each of the two major subsystems of AEROS, the National Emissions Data System (NEDS) and Storage and Retrieval of Aerometric Data (SAROAD), is described in detail in this paper. SAROAD is the established Federal data system for storing ambient air quality data from the air monitoring activities of State, local, and Federal agencies. NEDS, on the other hand, contains annual emissions and operating characteristics of individual emitters throughout the Nation. NEDS and SAROAD data are submitted regularly to EPA by all of the States in accordance with mandatory Federal reporting requirements.

It became apparent in 1973 that additional data were needed independently by the data bank users. For this reason, the concept of AEROS was expanded from an integrated NEDS/SAROAD data system to one encompassing other information systems such as test data, hazardous air pollutant sources, and computerized air pollution regulations. To date, the EPA air data systems have been used primarily by governmental agencies although the private sector is becoming more aware of the advantages in utilizing a common data base in conjunction with Government representatives. Thus, it is expected that the EPA air data systems will be used more frequently by private groups interested in influencing governmental decisions.

The Aerometric and Emissions Reporting System (AEROS) is comprised of input forms, programs, files, and reports established by EPA. These elements enable the EPA to collect, maintain, and report information describing air quality, emissions sources, and so forth. Although a great deal of the efforts and activities supporting AEROS are concerned primarily with the collection and maintenance of data, the primary purpose of AEROS is providing reports and computerized data on ambient air quality and emission sources. The input forms, procedures, programs, files, and reports are the basic structural elements of AEROS and, under the management of EPA, form a comprehensive system for collecting and reporting air quality and emissions data.

The purpose of AEROS is to provide hard data and basic information under the following requirements specified in the Clean Air Act:

- Evaluation of plans and strategies to meet national ambient air quality standards (including air pollution modeling)
- Evaluation of emissions and control equipment for the development of new source performance standards
- Support of hazardous pollutants investigations
- Determination of the status, projections, and trends of air pollution for reports and progress evaluation
- Studies on fuels, their usage, and availability.

AEROS is a general purpose data system which attempts to meet the general needs of a large number of users. It does not attempt to meet all of the needs of any of the users, and in fact, does not meet any of the requirements of certain potential users because of resource availability and certain other factors.

WHAT DATA ARE AVAILABLE?

Data Files

Data should be ordered in nonduplicative groups into many separate files tied together by identifiers. The groupings should be determined by the size, frequency of access, and available storage media (disk, tape). The same criteria should be applied to the decision on access speed (online-interactive, remote batch, etc.). Some duplication of files may be necessary if interactive access is considered.

AEROS

The AEROS system was developed over a long period by different groups before the concept of integrated data files and common data base management was accepted. Therefore, there is some duplication among the files which are not associated with the interactive system. Furthermore, as other program elements develop systems, there is a tendency to control files, which usually contain data duplicative of existing files, or to access existing files to create additional files. If permitted to continue unchecked, this practice will increase required storage space needlessly and will result in a string of unmanageable files in various stages of currency.

Research Data

Not all the data collected are intended for inclusion in AEROS. The primary purpose of AEROS is to serve general national needs; therefore, monitoring research data, short-term emission rates for models, or other data of short-term interest are not included. However, if these data are submitted in proper format by the collectors, they will not be rejected.

WHERE DID THE DATA COME FROM?

Basically, Federal regulations require the submittal of data in accordance with State Implementation Plans (SIP). Extensive reports are available indicating the types and sources of data received. The following statements are universally applicable to the major data banks.

Air Quality Data

Originally air quality data were obtained voluntarily from State and local agencies in exchange for provision of reports for State and local use. EPA research programs operated a 275-site network, performing all associated operations. The network was decreased in size and decentralized to Regional Offices (RO). Federal regulations require submittal of ambient data from networks as specified in SIP's.

Emissions Data

Initial efforts were centered on collecting emission inventories, in hard copy form, in the "first" 32 Air Quality Control Regions (AQCR) defined. States were provided funding to develop emission inventories in conjunction with SIP preparation. Summary data were to be submitted and detailed data kept on file. EPA attempted to collect all available data from States, local agencies, other Federal agencies, research projects, and other locations to create a base-line nationwide emission inventory. Currently, States are to submit emissions data in accordance with Federal regulations designed to maintain up-to-date information. Other EPA activities yielding emissions data are required to submit them to the banks in order to reduce duplication and redundancy.

HOW "GOOD" AND "COMPLETE" ARE THE DATA?

Quality assurance techniques for ambient monitoring have been developed by ORD; however, there has been no effort to address the matter of quality assurance in the emissions inventory area. The question concerning the data quality cannot be answered because there are no existing methods for quantitatively defining the assurance which a user can apply to the data. In general, there is more confidence in the quality of the ambient data than the emissions data.

The completeness of the ambient data may be considered from several perspectives:

- More sites are in operation than required by SIP's (for certain pollutants in certain areas); therefore, more data than necessary are being collected (about 300 percent over the required amount of data).
- State and local agencies manipulate the data and move monitoring sites, thereby causing gaps in the full picture of data and making statistical analysis difficult in some cases.
- Certain States and RO's are usually late in submitting the data and, therefore, are less complete.

The completeness of emissions data may be viewed according to the following:

- Some States and local agencies still do not have a usable emissions inventory.
- Certain States and local agencies have done an inadequate job of collecting the most basic information for calculating emissions; therefore, even though the sources have been identified, considerable information is missing.
- Some States have deliberately withheld data about selected sources; in some cases, RO's have known of the existence of sources but made no effort to include them in the system.

WHAT DO THE DATA BASE MANAGERS DO?

Considering the basic areas of responsibilities, i.e., data collection, communications, computer facility, data system, files, and software utilities, the data base managers perform the following functions in the two applicable areas of data system and files:

- Development of data system, including feasibility study, design, programing, testing, documentation, maintenance, and user surveys.
- Definition and creation of files, maintenance, add/delete/change actions, security, and auditing/anomaly investigations.
- Engineering (air pollution) necessary for calculations and statistics generation internal to the system.

Guidance is also provided by the data base managers with respect to data collection. No efforts are possible for interfacing data files with software utilities or for custom retrieval/analysis programing given the current level of resources. Data collection is by State and local agencies with overview of RO's. Communications, computer, and software utilities are the responsibility of the facility managers.

It is truly surprising to find a large number of EPA personnel in headquarters and regional offices who do not understand these divisions of responsibilities. On the other hand, perhaps they do not accept these divisions.

Most importantly, a conscious decision must be made by management concerning what the data systems are to accomplish and then to commit the necessary resources to support the desired level of accomplishment. The capabilities of ADP personnel must be thoroughly understood. Furthermore, it must be acknowledged that the crisis mode of operation can destroy a data base. Every operation, with few exceptions, must be viewed as a routine procedure; otherwise, the integrity of the system and the data base will be damaged. Management must also enforce the concept of nonredundant files so that valuable computer storage capability and operating time will not be wasted.

Finally, the philosophy of data base system operation must be uniform throughout the Agency. Otherwise, users who do not understand the differences among the many systems, will begin to use selected systems for uses for which they were not designed.

WHO ARE THE USERS OF THE DATA?

Users of AEROS may obtain data by the following methods:

- Use of existing batch and remote batch reports
- Use of existing interactive system reports
- Special requests to system/data base managers
- Writing of programs directly extracting data from defined files.

To date, the users have been identified as follows:

OAQPS, OAWM	50%
RO's	25%
Other EPA and public	25%

From another viewpoint, the users may be categorized according to the following:

- Persons or organizations for whom the system was designed and who, in general, have the majority of their needs met
- Persons or organizations who originally played a large part in the design of the system but currently do not use the system for some reason
- Persons or organizations not intended to be users who now insist on using the system to meet their needs.

There is another class of "users" which makes decisions without using available data, pretending that the data are nonexistent or "no good." Of course, it is easier to work without data, since data analysis is usually complex and time-consuming. However, this attitude cannot continue because sooner or later someone else will use the available data, come to a conclusion in conflict with the one previously made, and crisis/cover-up becomes the mode of operation.

Thus, it is very difficult to determine exactly who is a bona fide user and what the data needs are. This is further complicated by an inability to convince the users that the cost/benefit relationship indicates that all their needs cannot be completely met, but rather that a portion of all users' requirements can be met, given the resources available. Many users seem to be of the opinion that if every single requirement is not met in its entirety, the entire system is worthless. Unfortunately, these users also seem to have been able to convince others that this is a viable evaluation criteria.

WHAT ARE THE USES OF THE DATA?

The basic uses of the data are:

- Status reports on ambient concentrations emissions, fuel usage, and control equipment
- Assistance and guidance in the development of achievement, maintenance, and growth/non-degradation plans
- Evaluation of trends of air quality and emissions
- Forecasting of emissions and identification of potential problems.

It should be uniformly understood that data from a large computerized data bank should be used only for general purpose, demonstration, or management activities. If an important decision is to be made, the data should be used to identify a much smaller subset, and then the subset must be verified to assure freedom from error and degree of quality control.

DESCRIPTION AND CURRENT STATUS OF THE STRATEGIC ENVIRONMENTAL ASSESSMENT SYSTEM (SEAS)

By C. Lawrence

Since late 1972, the Office of Research and Development of the Environmental Protection Agency has been developing a system to assist in evaluating the impact of alternative regulatory actions and policies. The Strategic Environmental Assessment System (SEAS) is the result of this effort. The system combines analytical techniques, computer-based models, and expert opinions to forecast the probable impacts of environmental policies and strategies. Using the system, one can make alternative assumptions regarding economic growth, technological change in production processes and in pollution control, environmental policy, and regulatory actions. Reports generated are analyzed to ascertain the effects of the alternative assumptions.

The two primary functions of SEAS are:

- Prediction of pollutant loading on the environment from an assumed level of economic activity.
- Prediction of the abatement cost of that pollution if Federal pollution requirements are met.

This is usually a straightforward task. However, SEAS' job is complex because it must account for most forms of manmade pollution, and must do so at the national level. The job is further complicated by the necessity of projecting pollution rates into the future to analyze long-range effects of EPA policy and regulatory actions.

Long-range economic forecasting models are used to predict national economic activity. It is not necessary for the EPA to predict the economic course of the United States because most of the basic assumptions (gross national product, unemployment, population) can be taken from studies by the Department of Commerce. The model used to obtain the long-range projections and to drive the system is the INFORUM module. INFORUM, developed and maintained by the University of Maryland, provides a breakdown of U.S. economic activity. It computes activity levels for 185 industrial sectors and provides the following outputs for each year from 1971 through 1985:

- The 185 interindustry transactions
- Capital investment by 90 industries
- Material use by each of 28 construction categories
- Government purchases by 9 categories
- Inventory change
- Exports and imports
- Personal consumption expenditures

Although this approach provides a relatively detailed picture of the economy at the national level, it is not sufficiently detailed for residual, or pollution, forecasting. Therefore, the INFORUM sectors are disaggregated by the INSIDE module for those sectors where the rate of residual generation varies significantly over the range of products included in the sector, either as a result of technological changes or as a function of geographical location. Approximately 50 of the INFORUM sectors are disaggregated into 350 INSIDE subsectors. The output of the INFORUM and INSIDE modules provides the projected economic activity of the United States, distributed into 470 sectors/subsectors, over the years 1971 through 1985. (Sector outputs are in constant 1971 dollars, subsector outputs are in physical production units.)

These economic projections provide the basis for computing pollutant residuals by the RESGEN module. The sector/subsector economic outputs are converted to gross, or uncontrolled, residuals by application of gross residual coefficients (approximately 1,000 in number). Net residuals, after application of pollution control, and secondary residuals, created as a byproduct of gross residual primary control, are also computed. These residual forecasts are computed in units of annual tonnage of pollutant released to air, land, or water.

The SEAS ABATE module computes the cost of pollution abatement for approximately 500 abating sectors. These sectors may be either components or combinations of INFORUM sectors and INSIDE subsectors. Capital investment costs and operating/maintenance costs are computed for each abating sector based on: predicted treatment requirements; average plant size; and capital requirements for catchup, expansion, and replacement of pollution control equipment. The ABATE module also acts as the mechanism through which the economic impacts of pollution abatement are incorporated into the economy as a whole. The fundamental, or default, assumption of both the RESGEN and ABATE modules is that all relevant Federal pollution control requirements are complied with on schedule.

Each of the above-mentioned SEAS modules generates output files which are used as primary data sources by the remaining system modules. Other modules in SEAS include:

1. Solid Waste: this module estimates the annual tonnage at the national level of solid waste generated from all sources except pollution control (RESGEN module). Twelve materials categories (e.g., paper, glass, ferrous metals) and 20 product categories (e.g., newspaper, furniture, batteries) are considered. Estimates are computed for:

- Type of disposal facility–municipal or private
- Method of disposal–incineration, open dumping or landfill
- Disposal cost
- Levels of recycling and resultant secondary residuals

2. Regionalization: this module disaggregates national residuals estimates from the RESGEN module and national economic outputs from INFORUM and INSIDE into eight regional allocations including the following: states, Standard Metropolitan Statistical Areas (SMSA), major river basins, and minor river basins. Economic and residual shares are computed at the county level, and then reaggregrated to the desired regional allocation. The base year (1971) data from which the shares are computed is obtained from the best available source, with default values determined from an employment distribution survey. The projected change in these shares over time is computed from Department of Commerce projections (OBERS 2-digit SIC) for all industries except electric utilities, which use industry-published planning data.

3. Transportation: the transportation module forecasts the demand for automobile, bus, truck, railroad, and airline travel as vehicle miles traveled for both passenger and freight purposes. Based on these data, it estimates the total emissions produced by these sources. Emissions are computed in annual tonnage at the national level and may be disaggregated to the State and SMSA levels.

4. Energy: this module forecasts the energy demand resulting from the economic projections, by fuel type, for six user categories. The energy module is currently undergoing extensive revision.

5. Raw Materials: the STOCKS module estimates annual demand levels, relative price changes, capital investment, and import/export levels for 27 categories of raw resources.

6. Nonpoint Source Residuals: this module estimates the annual contribution of waterborne pollutants from agricultural and urban land use. It is currently in the test phase.

At the direction of Dr. Wilson K. Talley, EPA's Assistant Administrator for Research and Development, an ad hoc review panel was established by the Executive Committee of the EPA Science Advisory Board. The review panel was to assess the current status of SEAS and to make recommendations for its future development and application. This panel was headed by Nobel Laureate Wassily Leontief, who is currently at New York University. It recently completed its evaluation and made the following recommendations in its report:

- The SEAS system should be maintained in EPA and its utility enhanced by a carefully structured and independently reviewed program with the objectives of increasing the use of SEAS, developing a better data base, verifying results, and improving the structure.
- Encourage the use of SEAS and broaden the base for constructive criticism by a program to increase the visibility of SEAS to EPA Headquarters and Regions, regional and local officials and environmentalists, and other Federal agencies. Use for the Cost of Clean Air and Water Report is one example of a recommended use.

- Improve the residual coefficients in order of priority based on sensitivity tests to determine sectors of activity having the greatest impacts on residual generation and those sectors of high regulatory importance.

- Cooperate and share development costs with other agencies to develop modules and the corresponding data of mutual value to the agencies involved. Where these efforts are undertaken, the responsibility for developing models and collecting the data should rest in the other agencies so as not to detract from EPA efforts to improve SEAS for its own use.

A senior-management-level review team has been established to evaluate alternative ways of implementing the recommendations of the SAB review panel. The review team includes personnel from the Office of Planning and Evaluation as well as ORD. They will evaluate the alternatives within the practical limitations of existing budget and manpower. It is anticipated that final recommendations on resource levels and direction of developmental and operational activities will be forthcoming.

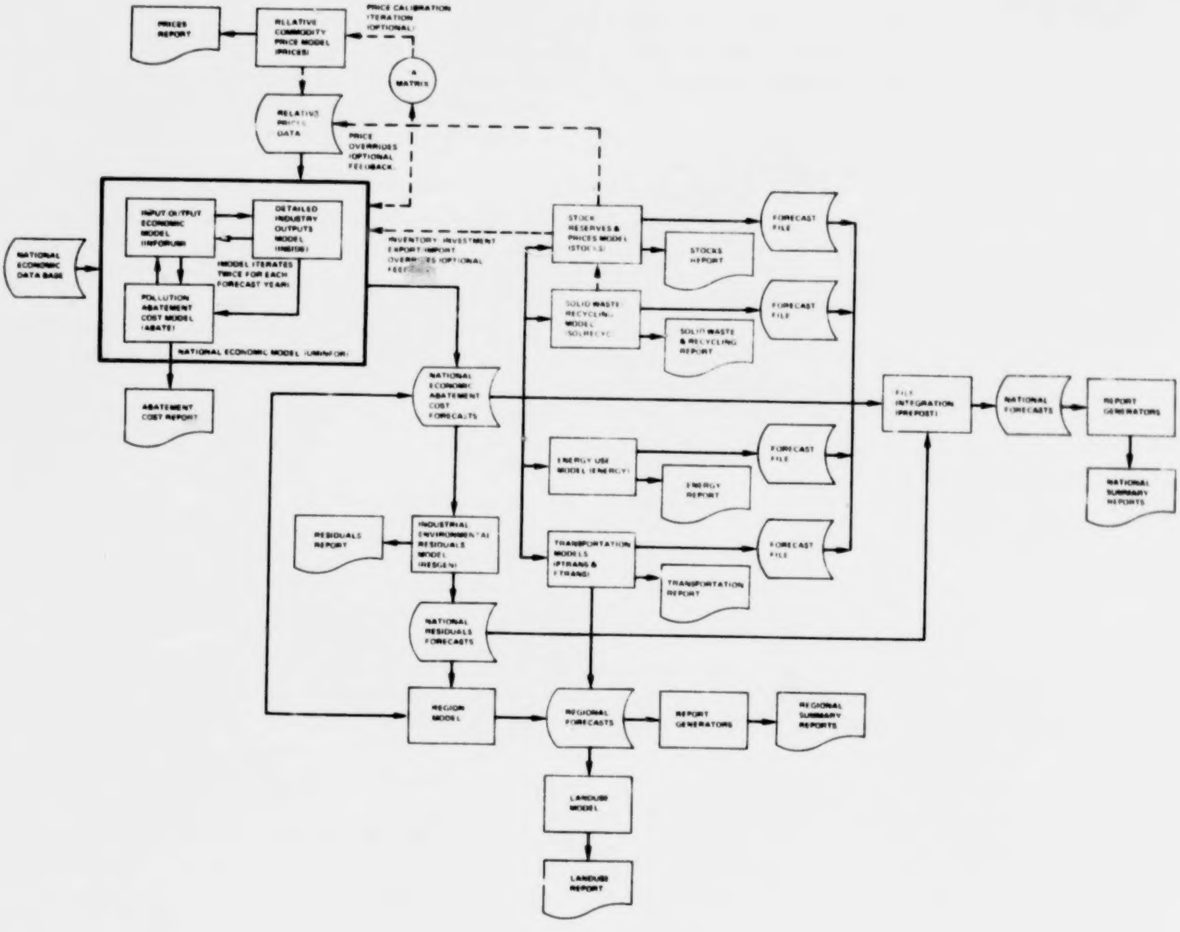

PRICES REPORT
RELATIVE COMMODITY PRICE MODEL (PRICES)
PRICE CALIBRATION ITERATION (OPTIONAL)
A MATRIX
RELATIVE PRICES DATA
PRICE OVERRIDES (OPTIONAL FEEDBACK)
NATIONAL ECONOMIC DATA BASE
INPUT OUTPUT ECONOMIC MODEL (INFORUM)
DETAILED INDUSTRY OUTPUTS MODEL (INSIDE)
(MODEL ITERATES TWICE FOR EACH FORECAST YEAR)
POLLUTION ABATEMENT COST MODEL (ABATE)
NATIONAL ECONOMIC MODEL (UMINFOR)
INVENTORY INVESTMENT EXPORT IMPORT OVERRIDES (OPTIONAL FEEDBACK)
ABATEMENT COST REPORT
STOCK RESERVES & PRICES MODEL (STOCKS)
STOCKS REPORT
FORECAST FILE
SOLID WASTE RECYCLING MODEL (SOLRECYC)
SOLID WASTE & RECYCLING REPORT
FORECAST FILE
NATIONAL ECONOMIC ABATEMENT COST FORECASTS
FILE INTEGRATION (PREPOST)
NATIONAL FORECASTS
REPORT GENERATORS
NATIONAL SUMMARY REPORTS
ENERGY USE MODEL (ENERGY)
ENERGY REPORT
FORECAST FILE
RESIDUALS REPORT
INDUSTRIAL ENVIRONMENTAL RESIDUALS MODEL (RESGEN)
TRANSPORTATION MODELS (PTRANS & FTRANS)
TRANSPORTATION REPORT
FORECAST FILE
NATIONAL RESIDUALS FORECASTS
REGION MODEL
REGIONAL FORECASTS
REPORT GENERATORS
REGIONAL SUMMARY REPORTS
LANDUSE MODEL
LANDUSE REPORT

Figure 1
SEAS Flow Diagram

IMPROVING THE UTILITY OF ENVIRONMENTAL SYSTEMS

By Donald Worley

Four random points are presented in this paper to help frame questions in the session which follows. These four points are in the form of questions:

- What is an environmental system?
- What are the environmental systems of EPA?
- Is our criticism valid?
- How have we determined our need?

Your view of the utility of environmental systems is dependent upon your particular backgrounds. The "old line" water quality people praise STORET as do those who have learned SAROAD. Those people coming to either system as a new user with preconceived ideas tend to criticize. Far too often their criticism is aimed at the wrong target.

To many people, SAROAD is a collection of computer programs written for IBM equipment and converted to Univac equipment. To others, SAROAD is 17 magnetic tapes of air quality data which may be used on the Univac 1110. In fact, neither of these opinions is correct. If SAROAD is viewed as an environmental system, then it includes much more than computer programs and data. Figure 1 is a simple graphic representation of SAROAD.

For a proper understanding of SAROAD, the following aspects of the system are important:

- The data is initially collected by State and local agencies
- The data is collected by a monitoring network that has been planned with Federal assistance and implemented by Federal money in many cases.
- The data is submitted through our Regional Offices to the National Air Data Branch of OAQPS
- Other aspects of this system.

Many Federal, State, and local policies are involved in this environmental system. Each component must carry its responsibility for the system to work effectively.

Figure 2 represents a view of some of the EPA environmental systems. Each system performs its mission uniquely, but each is separate. Little effort has been made to relate the methods or contents of these individual systems.

Few would champion a single EPA environmental system; however, there are existing techniques for relating common items within these systems. This relationship would not necessarily be a complex computer relationship, but rather could be a simple catalog of all environmental systems. At least, the catalog could detail places of reference for problems.

Criticism is a favorite American habit, and criticizing information systems is an easy thing to do. Much of our criticism is not unique to EPA systems. For 20 years, we in the computer field have been attempting to implement a system that will answer "any" question. This goal has not been achieved. Our current level is to answer the questions that we planned to answer.

Data management systems are the results of our search for the general solution. These systems are a large step forward from custom-designed programs, but their development is still ongoing. EPA information systems must rely upon these infant systems and some problems and limitations must, therefore, be expected.

Finally, we must consider our need for information systems. In the past, our group as well as other organizations have been easy prey for the salesman. The salesman brings a tool, and we try to find a way to use it. Unfortunately, the job we are doing does not need the tool in many cases. In fact, the tool that is unnecessary prevents us from obtaining the tool we need.

This discussion has been presented at an ORD workshop to focus on the fact that the users must help in defining the requirements for environmental systems.

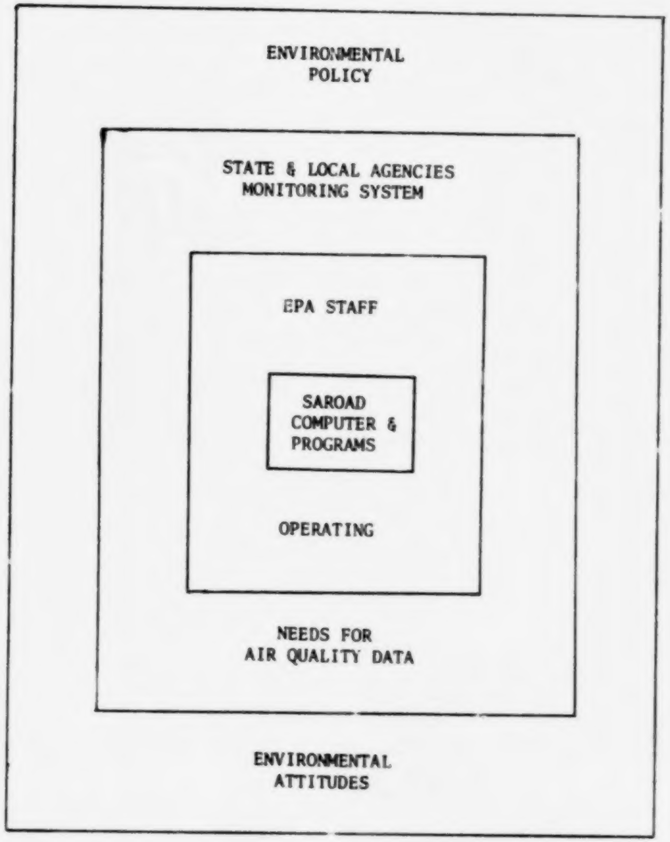

ENVIRONMENTAL POLICY
STATE & LOCAL AGENCIES MONITORING SYSTEM
EPA STAFF
SAROAD COMPUTER & PROGRAMS
OPERATING
NEEDS FOR AIR QUALITY DATA
ENVIRONMENTAL ATTITUDES

Figure 1
SAROAD

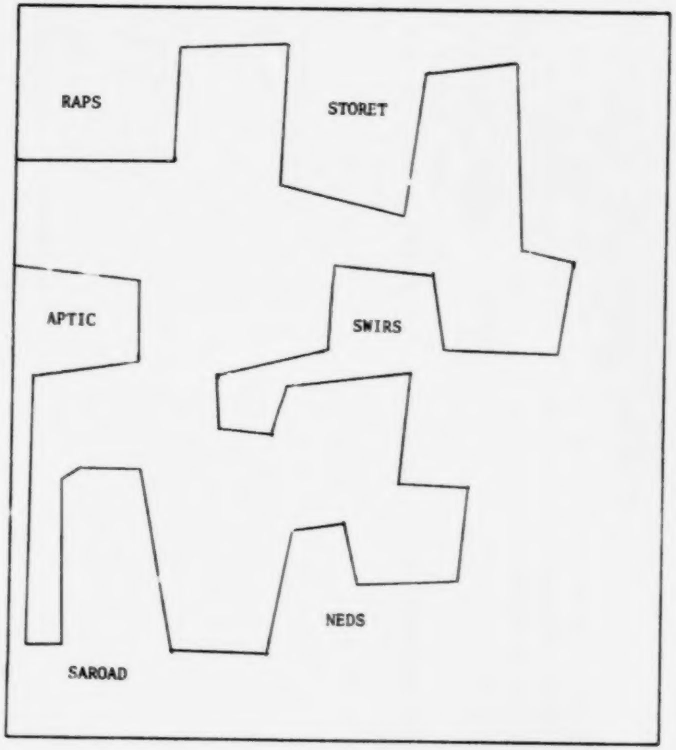

RAPS
STORET
APTIC
SWIRS
NEDS
SAROAD

Figure 2
The Information System Maze

SUMMARY OF DISCUSSION PERIOD - PANEL IV

The discussion from the session on utility of environmental data systems is presented below.

STORET Costs

First, it was stated that the annual costs to operate and maintain STORET are probably around $1 million. In the INFOMATICS study, which addressed the idea of moving STORET to the Univac at RTP, it was estimated that the move would cost between $7 and $11 million. The study used the word conversion, but the move might be better described as a reimplementation of a system design.

Survey Data Directory

It was pointed out that MIDSD's *Information Systems News* recently announced the publication of a loose-leaf directory of survey data studies. One panel member interpreted "survey data" to mean a *particular* data collection effort and results of that data collection as opposed to a *whole* system. It was suggested that a directory of surveys, related data files, and originators might be made available through a bibliographic data base system, but that a data base should not be created only because someone might possibly need the information.

STORET Routines

It was asked how to find out whether a program exists for a particular analysis of data in STORET. This information, it was pointed out, might be available through one's resident or Washington STORET contact. Normally, if a user produces a well-documented routine that could be of general interest, it will be made a standard part of the system. In addition, a list of routines people have used against the data base is maintained by the Washington STORET office. Information is also disseminated through meetings of active STORET users or potential users. This form of information interchange between the users is encouraged. Unfortunately, water quality does not have an ADP coordinator group set up like ORD, so the only focal point is at the Assistant Administrator level.

Functional Decisions

The management or functional decisions which demand immediate (online) access to these related data bases was discussed. A majority of the panel agreed that most decisions do not require the immediate response of an online system; generally the decision can wait for an overnight job to be run against a batch-supported system. However, one panel member felt that there are many instances when online access is needed. This allows a user to browse through the data and, possibly, to use interactive graphic packages against the data base. This is a tool to help make on-the-spot analysis of a selected subset of data.

Interactive Processing

The panel agreed that interactive processing is most appropriate when used for summaries, statistics, and general information, but not for detailed information. The discussion revealed a difference between a data base such as SAROAD or STORET and a bibliographic data base. Bibliographic data bases are cost effective online because of their very high usage. These systems also allow a user to develop a search heuristically and browse through the data. These characteristics are much more difficult or impossible to use and are time-consuming in a batch system. When appropriate, online systems allow a user to submit jobs to the batch operation and give him confidence that the job will be run overnight without errors in his job control language.

In a situation where a user wants to produce a subset of data from a large interactive system he is using, the panel felt data subsets should be produced in batch operations. The subset can then be searched online. The industry has shifted from encouraging all programing to be done over a terminal to suggesting that in many cases a batch operation might be more economical.

Future ADP Management

While the sessions have discussed the Agency mistakes that have been made in ADP, those present were interested in what input they are providing that gives hope for the future. The panel said these meetings have increased communication and have helped to break down the organizational barriers which apparently get in the way of data flow. They expressed the hope that these meetings would create enough excitement among attendees that they would return to their jobs and generate helpful input to management. Unfortunately, most input will have to go from the bottom of the organization to the decisionmakers at the top. Also, when the lower level rises to the top, they will be able to generate input. It was of concern to the panel that, although many branches of the Agency are represented at these meetings, if the attendees do not all choose one or two good methods to manage ADP, and go home and form completely divergent policies that are ultimately implemented, then nothing will have really been accomplished as a result of these meetings.

BIOSTORET

It was asked who will support BIOSTORET as it expands and acquires more users. BIOSTORET will be brought up as a pilot project in March 1976. It is assumed that OWHM will decide to support it at that time.

It was explained that BIOSTORET media is disk-dependent and this is because the hierarchical coding structure is a class order file which means it must be a direct access structure. The BIOSTORET file is not going to be a high data volume file in comparison with the water quality file. Additionally, BIOSTORET has excellent retrieval capabilities.

ADP System Users

How a data system manager determines who the users are, or should be, was discussed. Most members felt the manager must simply try to establish a class of user to fit the purpose of the data base and why it was accepted. At least the manager should establish the minimum and maximum levels of users. Another panel member felt the question reflects an underlying belief that data base systems should be judged by the number of users, which is erroneous. It should be judged by the user times his influence. If the President is the only user of a system and he finds the system useful, it is worthwhile. A positive value must be obtained when equating the cost and the gain. Once these positive results are obtained and the system is justified, other users are extra benefits. The manager does not need a large user community. If someone wishes to use a data base, it was generally agreed that the prospective user should go to the manager. The manager should have a set of guidelines on who may use an online system and, although the user may not be allowed online access, he will probably be provided with the data he requires from the system.

Members said that there has been much criticism among Agency groups about which groups are using particular data bases. The panel commented on what the Agency should be doing to stop this constant infighting and how it should rationally regard these systems.

The easiest method, the panel felt, is to make the user accountable for his utilization through the budget process. The individual manager should have the opportunity to use the most applicable system. Who should or should not use a system should not be dictated across the board.

ADP System Managers

Managers within EPA and those who evaluate EPA output should determine whether or not the reports and analyses they receive are adequate. It was not thought that any central committee or contractor could make the necessary assessment. Management should take steps to ensure that required data is collected and placed in the proper data base. If this is not done, the products of the data bases will be damaged.

Generally, the panel agreed that data system management should be placed in the line organization in the program offices. One member felt that data system management should be at the Assistant Administrator level but that users could be anywhere within the organization.

OPERATIONAL CHARACTERISTICS OF THE CHAMP DATA SYSTEM

By Marvin B. Hertz

Last year, a detailed description of the Community Health Air Monitoring Program (CHAMP system) was given at the workshop. The design of the system was basically complete at that time, and the Health Effects Research Lab was in the process of implementing the design.

The CHAMP system consists of a network of remote monitoring stations located across the country in coordination with concomitant epidemiologic studies. The two major requirements for the system were:

- To deliver high-quality data
- To deliver and handle large quantities of data.

To achieve these objectives, software has been developed (machine validation of the data) to connect aerometric and meteorologic data with system status information (i.e., instrument performance information). The validity, therefore, of each individual data point can be determined.

The remote station data acquisition system hardware is shown in Figure 1. Basically, the minicomputer in the remote station serves as an interface between the pollutant analyzers and associated system, magnetic tape data storage, the remote field service operator, and the telecommunications network. The data generated and recorded at the remotes and transmitted to central includes not only the actual meteorologic and pollutant sensor responses, but also associated analog signals and digital status signals (Figure 2). These signals supply information about the performance and status of each instrument. For example, if an instrument is switched from an ambient sampling mode to the calibration mode, a status bit is recorded which reflects this change.

The focal point of the CHAMP network is the central computer facility located at the Environmental Protection Agency, Research Triangle Park, North Carolina. The central controller for the CHAMP network is a dual processor system with a full complement of input, storage, and display peripherals. The heavy burden on processor time placed by the telecommunications and real-time processing of the large quantities of data justified the choice of a dual processor system. A PDP-11/40 with 40K of core was selected to perform the tasks associated with the management of the large data base generated by the network. The telecommunications and real-time processing tasks are handled by a PDP-11/05 computer with 16K of core. The two processors are interconnected by a Unibus window which takes advantage of the unified asynchronous data path architecture of the 11 system. The window allows each processor to address the core and peripherals on the other processor as if it were its own. In addition, the DEC memory management option was added to the PDP-11/40 to handle addressing above 32K in the 16-bit system. An extensive complement of peripherals including two 1.2M word cartridge type disks, three tape drives, an electrostatic printer-plotter, line printer, and CRT display were initially selected. The rapid retrieval requirements for large quantities of data necessitated the addition of a Telefile dual spindle, quad density, removable 20 surface pack disk system capable of storing 98M words. A block diagram of the Central Computer System is shown in Figure 3.

Although we could not have possibly contemplated all the problems that developed during the last year, the flexibility that was designed into the system enabled us to overcome most of the difficulties that resulted in the handling of the data. More important, however, the computer served as a valuable tool in solving many of the problems associated with system implementation.

The two major problems encountered were:

- Improper field testing and installation of the aerometric and meteorologic sensors prior to system startup
- Incomplete training of the field operators, especially in the area of total system design.

The following printouts demonstrate the various programs that were developed to allow the operation of each station to be followed at all times and to determine the validity of the data.

Figure 4 is a station map. It allows a remote data slot number (parameter) to be associated with a parameter mnemonic and an engineering unit mnemonic in all central operations. Since the system was limited by the number of possible data slots and the battery of instruments differed at the various stations, flexibility was thus added to the system.

Figure 5 is part of the validation criteria file. The primary parameter to be validated is indicated, and the concentration units are noted. The current calibration constants for the instrument are listed. These constants are updated as new calibration data are obtained from the remote data. The permissible delta about zero signifies the noise range of the instrument. The ten computer words listed under status bits comprise a validation map for status bits. A 1 in the top row of each word indicates that this bit must be checked. The 0 or 1 below this number indicates the valid condition for the instrument.

The minimum number of 5-minute averages required to make a valid hourly value is indicated, as well as the notation that no special software routine (handler) is required for this parameter.

Figure 6 is a continuation of the validation criteria files. The secondary parameters (analog signals to be tested) are listed as well as the upper and lower limits for correct instrument operation. The Validity Map Association gives the correspondence between a bit in the status word (carried along with every validated value) and the corresponding reason for invalidation.

The machine validation of the data is performed by Program VALDAT. Figure 7 is the first page of the output of VALDAT. The machine validated hourly values are presented by station and day for each parameter. The number after each parameter value denotes the number of valid 5-minute averages in that hourly value (12 maximum). The M, I, or B after each value signifies that the values not averaged are missing, invalid, or both missing and invalid respectively. The second page of VALDAT (Figure 8) lists the invalidity causes by parameter by hour. Figure 9 is the next part of VALDAT and lists the values of the invalid 5-minute averages which were invalidated by secondary parameters. The value of the secondary parameter is also listed. Figure 10, also part of VALDAT, lists the status of each 5-minute value by hour, by parameter. The characters -, O, I indicate that the noted 5-minute average is missing, valid, or invalid respectively. Figure 11 is a daily plot for a primary parameter for the station and day indicated. The # and I indicate valid and invalid data respectively. Since the line printer limited the number of points that could be plotted, basically every other 5-minute value is indicated.

VALDAT is reviewed by cognizant EPA personnel and the data edited to reflect station status that cannot be determined during the machine validation. For example, journal entries (Figure 12) and field logs supplied by the remote station operators are often helpful in validating the data. After the review of the data, changes are incorporated into the data base and a final "REVIEW" printout (Figure 13) is obtained. REVIEW is checked to see that the appropriate changes have been made. The data is then ready to be archived and used in the epidemiological analyses.

Although the system required considerable human intervention initially, the use of the software described in this paper has proven to be a valuable asset in the installation and maintenance of the stations. This will eventually lead to the need for only nominal human intervention (quality control spot checks) into the system.

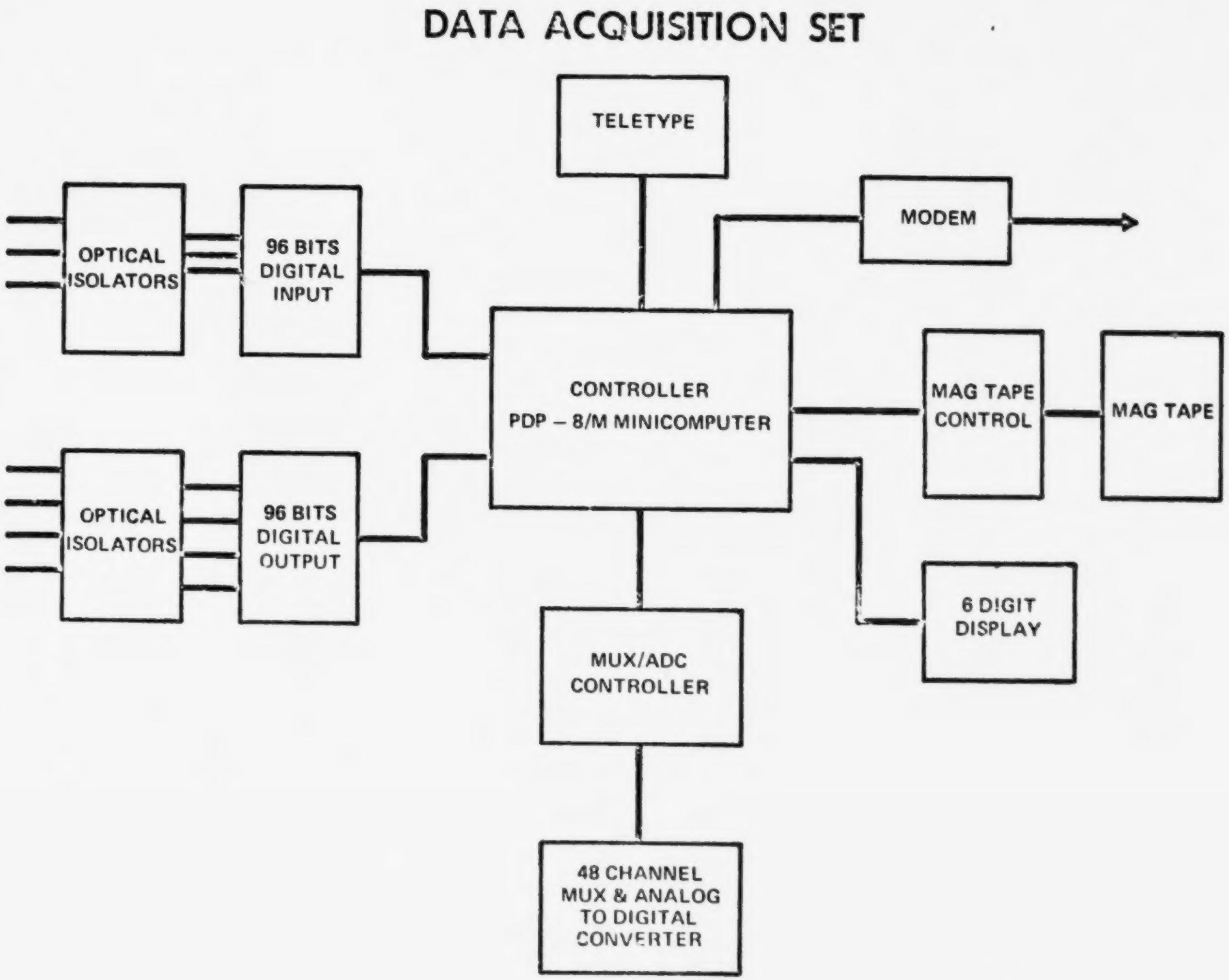

DATA ACQUISITION SET
TELETYPE
MODEM
OPTICAL ISOLATORS
96 BITS DIGITAL INPUT
CONTROLLER PDP – 8/M MINICOMPUTER
MAG TAPE CONTROL
MAG TAPE
OPTICAL ISOLATORS
96 BITS DIGITAL OUTPUT
6 DIGIT DISPLAY
MUX/ADC CONTROLLER
48 CHANNEL MUX & ANALOG TO DIGITAL CONVERTER

Figure 1
Remote Data Acquisition System

OZONE

- SAMPLE FLOW
- ETHYLENE FLOW
- POWER STATUS
- VALVE POSITIONS
- RANGE

} OPERATION CHECKS

- O_3 GENERATOR FLOW
- O_3 GENERATOR SETTING
- DILUTION AIR FLOW
- VALVE POSITIONS

} CALIBRATION CHECKS

Figure 2
Ozone Operational Parameters

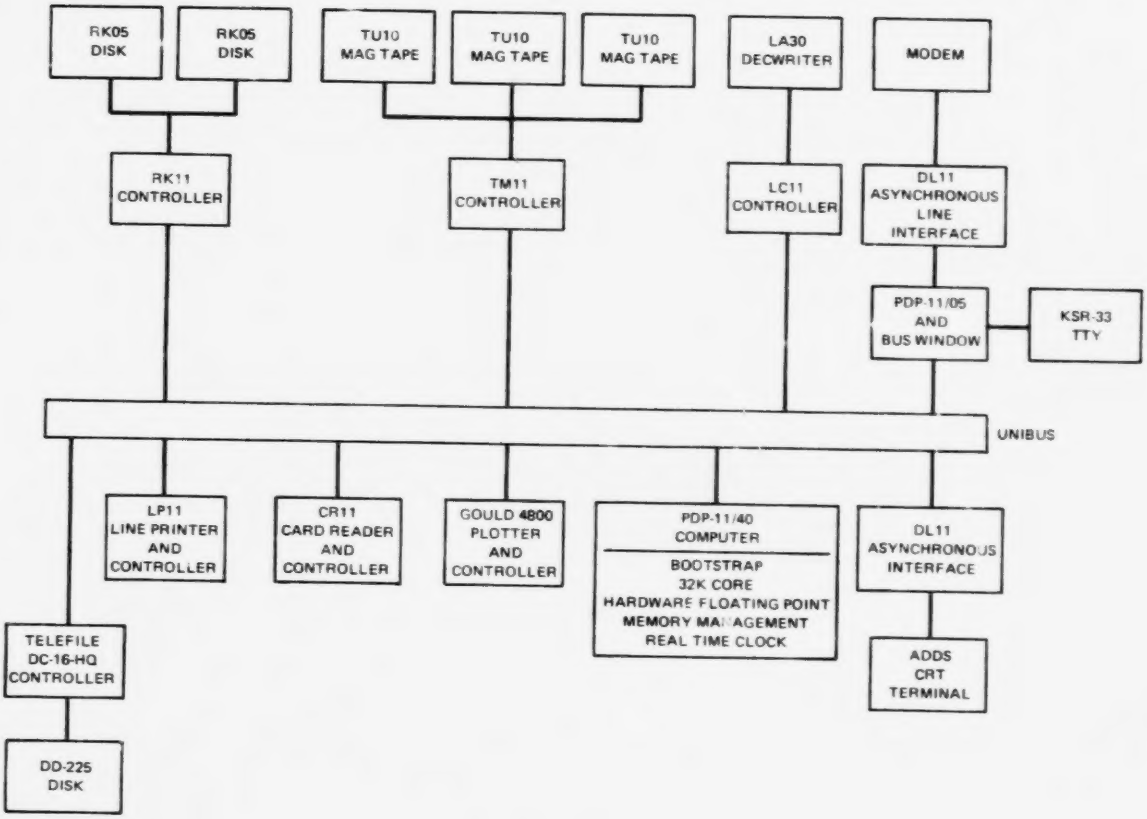

RK05 DISK
RK05 DISK
TU10 MAG TAPE
TU10 MAG TAPE
TU10 MAG TAPE
LA30 DECWRITER
MODEM
RK11 CONTROLLER
TM11 CONTROLLER
LC11 CONTROLLER
DL11 ASYNCHRONOUS LINE INTERFACE
PDP-11/05 AND BUS WINDOW
KSR-33 TTY
UNIBUS
LP11 LINE PRINTER AND CONTROLLER
CR11 CARD READER AND CONTROLLER
GOULD 4800 PLOTTER AND CONTROLLER
PDP-11/40 COMPUTER
BOOTSTRAP
32K CORE
HARDWARE FLOATING POINT
MEMORY MANAGEMENT
REAL TIME CLOCK
DL11 ASYNCHRONOUS INTERFACE
ADDS CRT TERMINAL
TELEFILE DC-16-HQ CONTROLLER
DD-225 DISK

Figure 3
Central Computer System Block Diagram

```
DK0:CR0032.MPP(2,2)

PARAMETER            MNEMONIC            EU MNEMONIC

     0                 WS                   K/HR
     1                 WD                   DEG
     2                 TOUT                 CENT
     3                 DEWP                 CENT
     4                 BP                   MMHG
     5                 TIN                  CENT
     6                 O3                   PPM
     7                 NO                   PPM
     8                 NO2                  PPM
     9                 NOX                  PPM
    10                 SO2                  PPM
    11                 0000                 0000
    12                 CO                   PPM
    13                 CH4                  PPM
    14                 THC                  PPM
    15                 MVFL                 L/M
    16                 WSPI                 L/M
    17                 WSPF                 L/M
    18                 PAN                  PPM
    19                 NMHC                 PPM
    20                 0000                 0000
    21                 PYRO                 VOLT
    22                 TPSL                 FARH
    23                 0000                 0000
    24                 0000                 0000
    25                 DAR1                 CC/M
    26                 DAR2                 CC/M
    27                 DAR3                 CC/M
    28                 GNO3                 %FS
    29                 HPAN                 CC/M
    30                 AFGC                 CC/M
    31                 TNOX                 CENT
    32                 WFFV                 VOLT
    33                 FETH                 CC/M
    34                 CALG                 CC/M
    35                 SFO3                 CC/M
    36                 SFNO                 CC/M
    37                 FO2                  CC/M
    38                 HSO2                 CC/M
    39                 SFSU                 CC/M
    40                 FCSN                 CC/M
    41                 0000                 0000
    42                 0000                 0000
    43                 0000                 0000
    44                 CALS                 UCCM
    45                 VAC                  TORR
    46                 LFL                  %LEL
    47                 HFGC                 CC/M
    48                 0000                 0020
    49                 0000                 0000
    50                 0000                 0000
    51                 0000                 0000
    52                 0000                 0000
    53                 0000                 0000
    54                 0000                 0000
    55                 0000                 0000
```

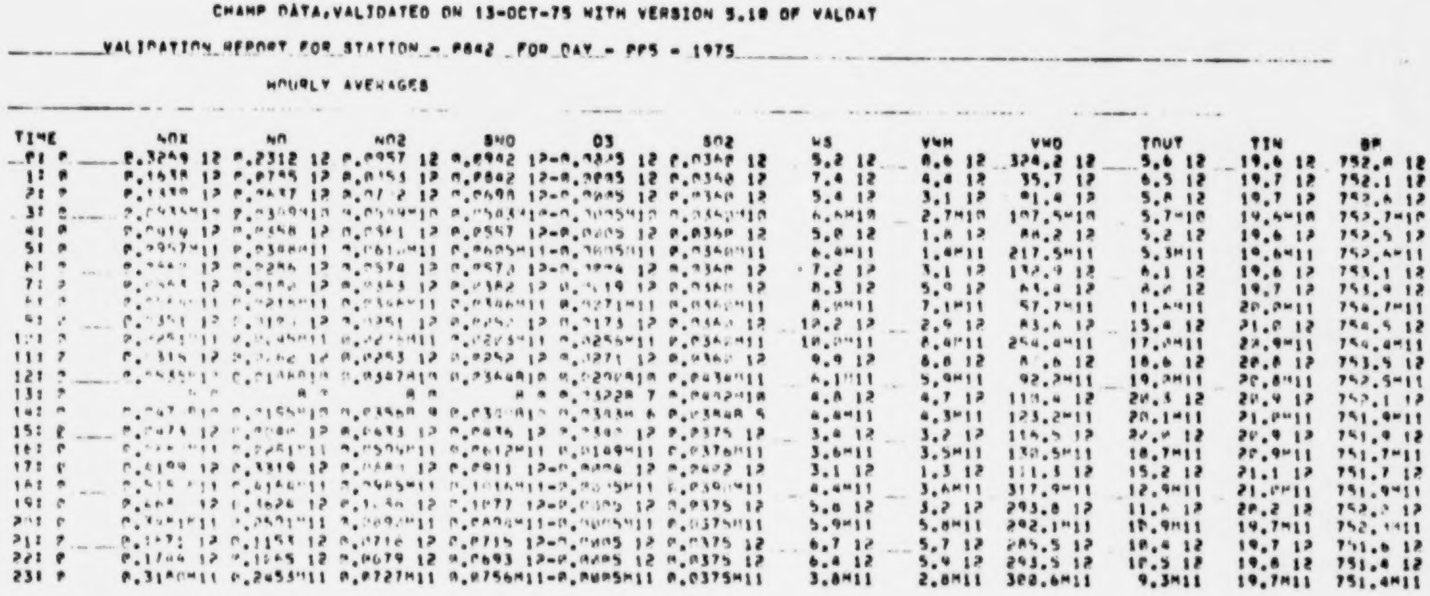
Figure 4
Station Identification Map

```
                                LISTING OF FILE DK1:VC0041.001
PRIMARY PARAMETER: O3   IN PPM

CURRENT CALIBRATION CONSTANTS ARE:
A=   -0.0071   B=    0.1057 C=    0.0008  MOLF =   1

PERMISSIBLE DELTA ABOUT ZERO =  0.010 PPM
REPORT FORMAT - F7.4

                                  ************
                           STATUS BITS

WORD/BIT         0    1    2    3    4    5    6    7    8    9   10   11
WORD  0 EX       0    0    0    0    0    0    0    0    0    0    0    0
VAL
WORD  1 EX       0    0    0    0    0    0    0    0    0    0    0    0
VAL
WORD  2 EX       0    0    0    0    0    0    0    0    0    0    0    0
VAL
WORD  3 EX       0    0    0    0    0    0    0    0    0    0    0    0
VAL
WORD  4 EX       0    0    0    0    0    0    0    0    0    0    0    0
VAL
WORD  5 EX       0    0    0    0    0    0    0    0    0    0    0    0
VAL
WORD  6 EX       0    0    0    0    0    0    0    0    0    0    0    0
VAL
WORD  7 EX       0    0    0    0    0    0    0    0    0    0    0    0
VAL
WORD  8 EX       0    0    0    0    0    0    0    0    0    1    0    0
VAL                                                         0
WORD  9 EX       0    0    0    0    0    0    0    0    0    0    0    0
VAL

MINIMUM NO. OF POINTS TO MAKE A VALID HOUR AVG = 9

NO SPECIAL HANDLER FOR THIS PRIMARY PARAMETER
```

Figure 5
Validation Criteria File I

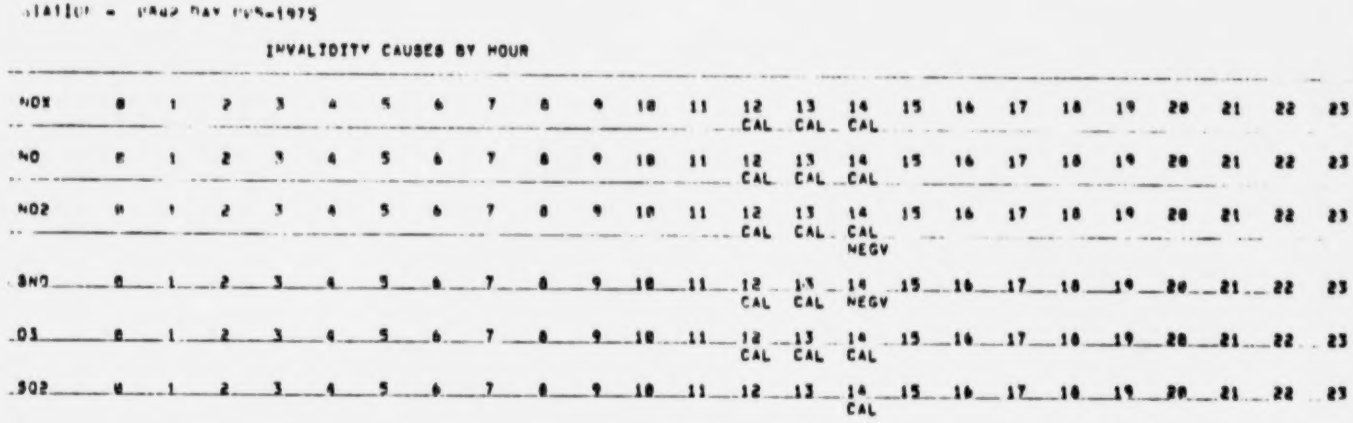
```
NUMBER OF ASSOCIATED SECONDARY PARAMETERS =  2

SEC PARAM=  1    FETH
A=     1.0920B=     4.509C=     0.161
LOW LIMIT =   10.000             HIGH LIMIT =    32.000

SEC PARAM=  2    SFO3
A=  105.170B=   393.300C=   -2.075
LOW LIMIT =  920.000             HIGH LIMIT = 1120.000

                    VALIDITY MAP ASSOCIATION

BIT    WORD/SPAR     BIT        MNEMONIC
  0       S 1                     FETH
  1       S 2                     SFO3
 25       W 0         8           PWR
 35       W 0         9           CAL
 45       W 1         6
 55       W 1         8
 65       W 1         9
 75       W 6         4
 85       W 0         0
 95       W 0         0
105       W 0         0
115       W 0         0
125       W 0         0
135       W 0         0
```

Figure 6
Validation Criteria File II



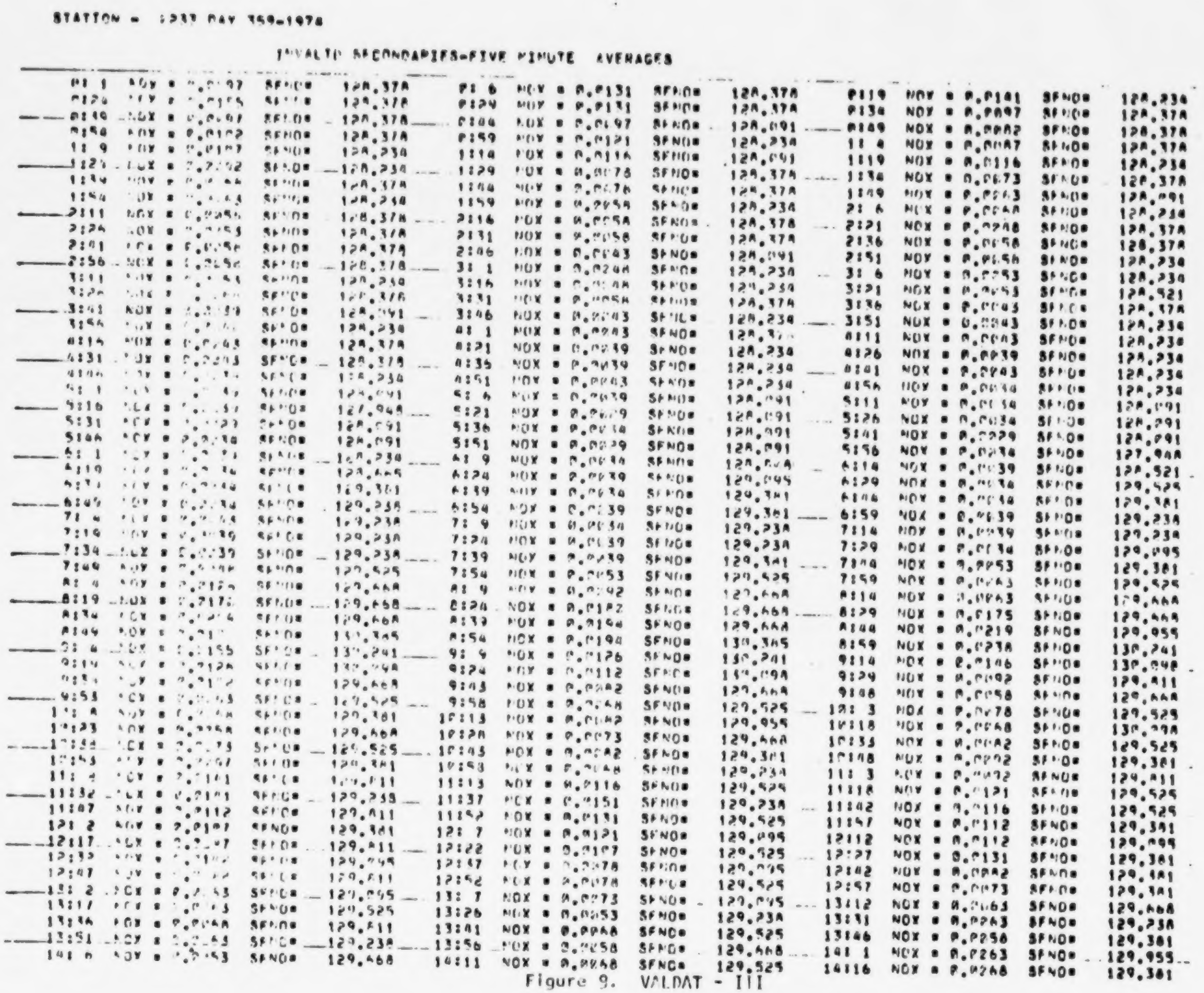
```
                    CHAMP DATA,VALIDATED ON 13-OCT-75 WITH VERSION 5.10 OF VALDAT
______VALIDATION REPORT FOR STATION - 0842  FOR DAY - 005 - 1975
                    HOURLY AVERAGES

TIME      NOX        NO         NO2        SNO        O3         SO2         WS       VWM      VWD        TOUT      TIN       BP
 0: 0   0.3249 12 0.2312 12 0.0957 12 0.0902 12-0.0025 12 0.0360 12     5.2 12    0.6 12  324.2 12    5.6 12   19.6 12  752.0 12
 1: 0   0.1638 12 0.0795 12 0.0153 12 0.0802 12-0.0035 12 0.0360 12     7.4 12    4.4 12   35.7 12    6.5 12   19.7 12  752.1 12
 2: 0   0.1130 12 0.0437 12 0.0712 12 0.0690 12-0.0085 12 0.0360 12     5.4 12    3.1 12   81.4 12    5.0 12   19.7 12  752.6 12
 3: 0   0.0935M10 0.0309M10 0.0599M10 0.0583M10-0.0005M10 0.0360M10     6.6M10    2.7M10  107.5M10    5.7M10   19.6M10  752.7M10
 4: 0   0.0919 12 0.0358 12 0.0561 12 0.0557 12-0.0005 12 0.0360 12     5.0 12    1.8 12   88.2 12    5.2 12   19.6 12  752.5 12
 5: 0   0.0957M11 0.0300M11 0.0610M11 0.0605M11-0.0005M11 0.0360M11     6.4M11    1.4M11  217.5M11    5.3M11   19.6M11  752.6M11
 6: 0   0.0860 12 0.0286 12 0.0574 12 0.0573 12-0.0006 12 0.0360 12     7.2 12    3.1 12  132.9 12    6.1 12   19.6 12  753.1 12
 7: 0   0.0565 12 0.0182 12 0.0383 12 0.0382 12 0.0119 12 0.0360 12     8.3 12    5.9 12   63.4 12    8.0 12   19.7 12  753.9 12
 8: 0   0.[illegible]M11 0.0216M11 0.0364M11 0.0366M11 0.0271M11 0.0360M11     8.0M11    7.1M11   57.7M11   11.6M11   20.0M11  754.7M11
 9: 0   0.0351 12 0.0100 12 0.0251 12 0.0252 12 0.0173 12 0.0360 12    12.2 12    2.9 12   83.6 12   15.4 12   21.0 12  754.5 12
10: 0   0.0281M11 0.0056M11 0.0225M11 0.0225M11 0.0256M11 0.0360M11    10.0M11    8.0M11  254.0M11   17.0M11   22.9M11  754.4M11
11: 0   0.0315 12 0.0062 12 0.0253 12 0.0252 12 0.0271 12 0.0360 12     9.9 12    6.8 12   87.6 12   18.6 12   20.8 12  753.5 12
12: 0   0.0555M10 0.0180M10 0.0347M10 0.0364M10 0.0200M10 0.0430M11     6.1M11    5.9M11   92.2M11   19.2M11   20.8M11  752.5M11
13: 0        0 0        0 0        0 0        0 0 0.1322  7 0.0402M10     6.8 12    4.7 12  119.4 12   20.3 12   20.9 12  752.1 12
14: 0   0.0470M10 0.0155M10 0.0356 0 0.0320M10 0.0303  6 0.0384  5     6.4M11    4.3M11  123.2M11   20.1M11   21.0M11  751.9M11
15: 0   0.0474 12 0.0000 12 0.0433 12 0.0436 12 0.0340 12 0.0375 12     3.4 12    3.2 12  116.5 12   22.0 12   20.9 12  751.9 12
16: 0   0.[illegible]M11 0.2261M11 0.0509M11 0.0612M11 0.0109M11 0.0376M11     3.6M11    3.5M11  130.5M11   18.7M11   20.9M11  751.7M11
17: 0   0.4109 12 0.3319 12 0.0801 12 0.0911 12-0.0006 12 0.0402 12     3.1 12    1.3 12  111.3 12   15.2 12   21.1 12  751.7 12
18: 0   0.5150M11 0.4160M11 0.0985M11 0.1016M11-0.0015M11 0.0390M11     4.4M11    3.6M11  317.9M11   12.9M11   21.0M11  751.9M11
19: 0   0.6006 12 0.4626 12 0.1380 12 0.1077 12-0.0005 12 0.0375 12     5.8 12    3.2 12  293.8 12   11.6 12   20.2 12  752.2 12
20: 0   0.3401M11 0.2501M11 0.0892M11 0.0894M11-0.0005M11 0.0375M11     5.9M11    5.0M11  292.1M11   10.9M11   19.7M11  752.1M11
21: 0   0.1872 12 0.1153 12 0.0716 12 0.0715 12-0.0005 12 0.0375 12     6.7 12    5.7 12  285.5 12   10.4 12   19.7 12  751.6 12
22: 0   0.1744 12 0.1065 12 0.0679 12 0.0693 12-0.0005 12 0.0375 12     6.4 12    5.4 12  293.5 12   10.5 12   19.8 12  751.4 12
23: 0   0.3180M11 0.2453M11 0.0727M11 0.0756M11-0.0005M11 0.0375M11     3.8M11    2.0M11  308.6M11    9.3M11   19.7M11  751.4M11
```

Figure 7
VALDAT-I



```
STATION - 0842 DAY 005-1975

              INVALIDITY CAUSES BY HOUR

NOX     0   1   2   3   4   5   6   7   8   9   10   11   12   13   14   15   16   17   18   19   20   21   22   23
                                                         CAL  CAL  CAL
NO      0   1   2   3   4   5   6   7   8   9   10   11   12   13   14   15   16   17   18   19   20   21   22   23
                                                         CAL  CAL  CAL
NO2     0   1   2   3   4   5   6   7   8   9   10   11   12   13   14   15   16   17   18   19   20   21   22   23
                                                         CAL  CAL  CAL
                                                                   NEGV
SNO     0   1   2   3   4   5   6   7   8   9   10   11   12   13   14   15   16   17   18   19   20   21   22   23
                                                         CAL  CAL  NEGV
O3      0   1   2   3   4   5   6   7   8   9   10   11   12   13   14   15   16   17   18   19   20   21   22   23
                                                         CAL  CAL  CAL
SO2     0   1   2   3   4   5   6   7   8   9   10   11   12   13   14   15   16   17   18   19   20   21   22   23
                                                                   CAL
```

Figure 8
VALDAT-II



```
STATION =  1237 DAY 359-1978

                    INVALID SECONDARIES-FIVE MINUTE  AVERAGES

 0: 1  NOX = 0.0197      SFNO=  128.378    0: 6  NOX = 0.0131  SFNO=  128.378    0:19  NOX = 0.0141  SFNO=  128.238
 0:24  NOX = 0.0155      SFNO=  128.378    0:29  NOX = 0.0131  SFNO=  128.378    0:34  NOX = 0.0097  SFNO=  128.378
 0:39  NOX = [illegible] SFNO=  128.378    0:44  NOX = 0.0097  SFNO=  128.091    0:49  NOX = 0.0082  SFNO=  128.378
 0:54  NOX = 0.0102      SFNO=  128.378    0:59  NOX = 0.0121  SFNO=  128.238    1: 4  NOX = 0.0087  SFNO=  128.378
 1: 9  NOX = 0.0107      SFNO=  128.238    1:14  NOX = 0.0116  SFNO=  128.091    1:19  NOX = 0.0116  SFNO=  128.238
 1:24  NOX = [illegible] SFNO=  128.238    1:29  NOX = 0.0078  SFNO=  128.378    1:34  NOX = 0.0073  SFNO=  128.378
 1:39  NOX = [illegible] SFNO=  128.378    1:44  NOX = 0.0078  SFNO=  128.378    1:49  NOX = 0.0063  SFNO=  128.091
 1:54  NOX = [illegible] SFNO=  128.238    1:59  NOX = 0.0058  SFNO=  128.238    2: 6  NOX = 0.0068  SFNO=  128.238
 2:11  NOX = [illegible] SFNO=  128.378    2:16  NOX = 0.0058  SFNO=  128.378    2:21  NOX = 0.0048  SFNO=  128.378
 2:26  NOX = [illegible] SFNO=  128.378    2:31  NOX = 0.0058  SFNO=  128.378    2:36  NOX = 0.0058  SFNO=  128.378
 2:41  NOX = 0.0058      SFNO=  128.378    2:46  NOX = 0.0043  SFNO=  128.091    2:51  NOX = 0.0058  SFNO=  128.238
 2:56  NOX = [illegible] SFNO=  128.378    3: 1  NOX = 0.0048  SFNO=  128.238    3: 6  NOX = 0.0053  SFNO=  128.238
 3:11  NOX = [illegible] SFNO=  128.238    3:16  NOX = 0.0048  SFNO=  128.238    3:21  NOX = 0.0053  SFNO=  128.521
 3:26  NOX = [illegible] SFNO=  128.378    3:31  NOX = 0.0058  SFNO=  128.378    3:36  NOX = 0.0043  SFNO=  128.378
 3:41  NOX = [illegible] SFNO=  128.091    3:46  NOX = 0.0043  SFNO=  128.238    3:51  NOX = 0.0043  SFNO=  128.238
 3:56  NOX = [illegible] SFNO=  128.238    4: 1  NOX = 0.0043  SFNO=  128.378    4:11  NOX = 0.0043  SFNO=  128.238
 4:16  NOX = 0.0043      SFNO=  128.378    4:21  NOX = 0.0039  SFNO=  128.238    4:26  NOX = 0.0039  SFNO=  128.238
 4:31  NOX = [illegible] SFNO=  128.378    4:35  NOX = 0.0039  SFNO=  128.238    4:41  NOX = 0.0043  SFNO=  128.238
 4:46  NOX = [illegible] SFNO=  128.238    4:51  NOX = 0.0043  SFNO=  128.238    4:56  NOX = 0.0034  SFNO=  128.238
 5: 1  NOX = [illegible] SFNO=  128.091    5: 6  NOX = 0.0039  SFNO=  128.091    5:11  NOX = 0.0034  SFNO=  128.091
 5:16  NOX = [illegible] SFNO=  127.948    5:21  NOX = 0.0029  SFNO=  128.091    5:26  NOX = 0.0034  SFNO=  128.091
 5:31  NOX = [illegible] SFNO=  128.091    5:36  NOX = 0.0034  SFNO=  128.091    5:41  NOX = 0.0029  SFNO=  128.091
 5:46  NOX = [illegible] SFNO=  128.091    5:51  NOX = 0.0029  SFNO=  128.091    5:56  NOX = 0.0034  SFNO=  127.948
 6: 1  NOX = [illegible] SFNO=  128.238    6: 9  NOX = 0.0034  SFNO=  [illegible]    6:14  NOX = 0.0039  SFNO=  128.521
 6:19  NOX = [illegible] SFNO=  128.665    6:24  NOX = 0.0039  SFNO=  129.095    6:29  NOX = 0.0034  SFNO=  129.525
 6:34  NOX = [illegible] SFNO=  129.381    6:39  NOX = 0.0034  SFNO=  129.381    6:44  NOX = 0.0034  SFNO=  129.381
 6:49  NOX = [illegible] SFNO=  129.238    6:54  NOX = 0.0039  SFNO=  129.381    6:59  NOX = 0.0039  SFNO=  129.238
 7: 4  NOX = [illegible] SFNO=  129.238    7: 9  NOX = 0.0034  SFNO=  129.238    7:14  NOX = 0.0039  SFNO=  129.238
 7:19  NOX = [illegible] SFNO=  129.238    7:24  NOX = 0.0039  SFNO=  129.238    7:29  NOX = 0.0034  SFNO=  129.095
 7:34  NOX = [illegible] SFNO=  129.238    7:39  NOX = 0.0039  SFNO=  129.381    7:44  NOX = 0.0053  SFNO=  129.381
 7:49  NOX = [illegible] SFNO=  129.525    7:54  NOX = 0.0053  SFNO=  129.525    7:59  NOX = 0.0063  SFNO=  129.525
 8: 4  NOX = [illegible] SFNO=  129.668    8: 9  NOX = 0.0092  SFNO=  129.668    8:14  NOX = 0.0063  SFNO=  129.668
 8:19  NOX = [illegible] SFNO=  129.668    8:24  NOX = 0.0182  SFNO=  129.668    8:29  NOX = 0.0175  SFNO=  129.668
 8:34  NOX = [illegible] SFNO=  129.668    8:39  NOX = 0.0194  SFNO=  129.668    8:44  NOX = 0.0219  SFNO=  129.955
 8:49  NOX = [illegible] SFNO=  130.385    8:54  NOX = 0.0194  SFNO=  130.385    8:59  NOX = 0.0238  SFNO=  130.241
 9: 4  NOX = 0.0155      SFNO=  130.241    9: 9  NOX = 0.0126  SFNO=  130.241    9:14  NOX = 0.0146  SFNO=  130.098
 9:19  NOX = 0.0126      SFNO=  130.098    9:24  NOX = 0.0112  SFNO=  130.098    9:29  NOX = 0.0092  SFNO=  129.811
 9:34  NOX = [illegible] SFNO=  129.668    9:43  NOX = 0.0082  SFNO=  129.668    9:48  NOX = 0.0058  SFNO=  129.668
 9:53  NOX = [illegible] SFNO=  129.525    9:58  NOX = 0.0068  SFNO=  129.525   10: 3  NOX = 0.0078  SFNO=  129.525
10: 8  NOX = [illegible] SFNO=  129.381   10:13  NOX = 0.0082  SFNO=  129.955   10:18  NOX = 0.0068  SFNO=  130.098
10:23  NOX = [illegible] SFNO=  129.668   10:28  NOX = 0.0073  SFNO=  129.668   10:33  NOX = 0.0082  SFNO=  129.525
10:38  NOX = [illegible] SFNO=  129.525   10:43  NOX = 0.0082  SFNO=  129.381   10:48  NOX = 0.0092  SFNO=  129.381
10:53  NOX = [illegible] SFNO=  129.381   10:58  NOX = 0.0048  SFNO=  129.238   11: 3  NOX = 0.0072  SFNO=  129.811
11: 4  NOX = [illegible] SFNO=  129.811   11:13  NOX = 0.0116  SFNO=  129.525   11:18  NOX = 0.0121  SFNO=  129.525
11:32  NOX = [illegible] SFNO=  129.238   11:37  NOX = 0.0151  SFNO=  129.238   11:42  NOX = 0.0116  SFNO=  129.525
11:47  NOX = 0.0112      SFNO=  129.811   11:52  NOX = 0.0131  SFNO=  129.525   11:57  NOX = 0.0112  SFNO=  129.381
12: 2  NOX = 0.0107      SFNO=  129.381   12: 7  NOX = 0.0121  SFNO=  129.095   12:12  NOX = 0.0112  SFNO=  129.095
12:17  NOX = [illegible] SFNO=  129.811   12:22  NOX = 0.0107  SFNO=  129.525   12:27  NOX = 0.0131  SFNO=  129.381
12:32  NOX = [illegible] SFNO=  129.095   12:37  NOX = 0.0078  SFNO=  129.095   12:42  NOX = 0.0082  SFNO=  129.381
12:47  NOX = [illegible] SFNO=  129.811   12:52  NOX = 0.0078  SFNO=  129.525   12:57  NOX = 0.0073  SFNO=  129.381
13: 2  NOX = [illegible] SFNO=  129.095   13: 7  NOX = 0.0073  SFNO=  129.095   13:12  NOX = 0.0063  SFNO=  129.668
13:17  NOX = [illegible] SFNO=  129.525   13:26  NOX = 0.0053  SFNO=  129.238   13:31  NOX = 0.0263  SFNO=  129.238
13:36  NOX = 0.0048      SFNO=  129.811   13:41  NOX = 0.0068  SFNO=  129.525   13:46  NOX = 0.0256  SFNO=  129.381
13:51  NOX = [illegible] SFNO=  129.238   13:56  NOX = 0.0058  SFNO=  129.668   14: 1  NOX = 0.0263  SFNO=  129.955
14: 6  NOX = [illegible] SFNO=  129.668   14:11  NOX = 0.0068  SFNO=  129.525   14:16  NOX = 0.0268  SFNO=  129.381
```

Figure 9. VALDAT - III

Figure 9
VALDAT-III

```
STATION =  0842 DAY 005-1975

                                   ****DATA REVIEW****
-----------------------------------------------------------------------------------------------------
  NOX              0            1            2            3            4            5            6            7
-----------000000000000 000000000000 000000000000 0000--000000 000000000000 00000-000000 000000000000 000000000000
                   8            9           10           11           12           13           14           15
-----------0000000000-0 000000000000 0000000000-0 000000000000 0000000000-I I----------- I000000000-0 000000000000
                  16           17           18           19           20           21           22           23
-----------00000000000- 000000000000 0000000000-0 000000000000 00000000000- 000000000000 000000000000 0000000-0000
  NO               0            1            2            3            4            5            6            7
-----------000000000000 000000000000 000000000000 0000--000000 000000000000 00000-000000 000000000000 000000000000
                   8            9           10           11           12           13           14           15
-----------0000000000-0 000000000000 0000000000-0 000000000000 0000000000-I I----------- I000000000-0 000000000000
                  16           17           18           19           20           21           22           23
-----------00000000000- 000000000000 0000000000-0 000000000000 00000000000- 000000000000 000000000000 0000000-0000
  NO2              0            1            2            3            4            5            6            7
-----------000000000000 000000000000 000000000000 0000--000000 000000000000 00000-000000 000000000000 000000000000
                   8            9           10           11           12           13           14           15
-----------0000000000-0 000000000000 0000000000-0 000000000000 0000000000-I I----------- I0000I0000-0 000000000000
                  16           17           18           19           20           21           22           23
-----------00000000000- 000000000000 0000000000-0 000000000000 00000000000- 000000000000 000000000000 0000000-0000
  SNO              0            1            2            3            4            5            6            7
-----------000000000000 000000000000 000000000000 0000--000000 000000000000 00000-000000 000000000000 000000000000
                   8            9           10           11           12           13           14           15
-----------0000000000-0 000000000000 0000000000-0 000000000000 0000000000-I I----------- 00000I0000-0 000000000000
                  16           17           18           19           20           21           22           23
-----------00000000000- 000000000000 0000000000-0 000000000000 00000000000- 000000000000 000000000000 0000000-0000
  O3               0            1            2            3            4            5            6            7
-----------000000000000 000000000000 000000000000 0000--000000 000000000000 00000-000000 000000000000 000000000000
                   8            9           10           11           12           13           14           15
-----------0000000000-0 000000000000 0000000000-0 000000000000 0000000000-I I--I0000000I ---I00000--0 000000000000
                  16           17           18           19           20           21           22           23
-----------00000000000- 000000000000 0000000000-0 000000000000 00000000000- 000000000000 000000000000 0000000-0000
  SO2              0            1            2            3            4            5            6            7
-----------000000000000 000000000000 000000000000 0000--000000 000000000000 00000-000000 000000000000 000000000000
                   8            9           10           11           12           13           14           15
-----------0000000000-0 000000000000 0000000000-0 000000000000 0000000000-0 0--000000000 000-----I0-0 000000000000
                  16           17           18           19           20           21           22           23
-----------00000000000- 000000000000 0000000000-0 000000000000 00000000000- 000000000000 000000000000 0000000-0000
  WS               0            1            2            3            4            5            6            7
-----------000000000000 000000000000 000000000000 0000--000000 000000000000 00000-000000 000000000000 000000000000
```

Figure 10
VALDAT-IV

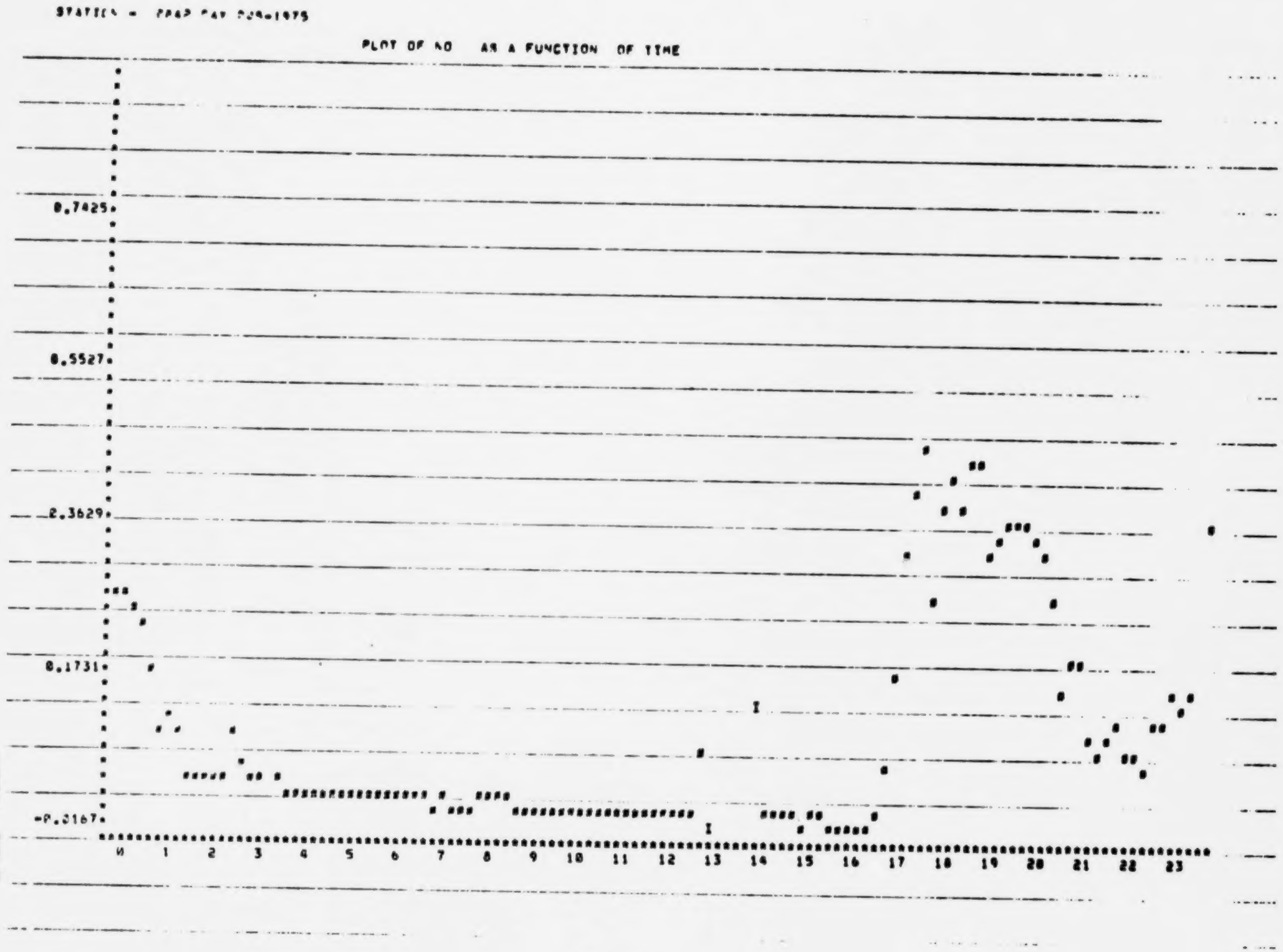
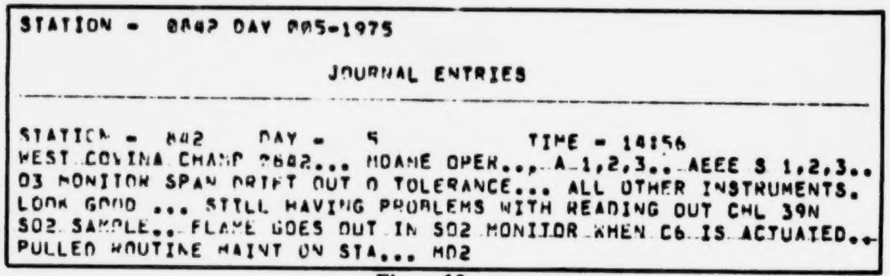
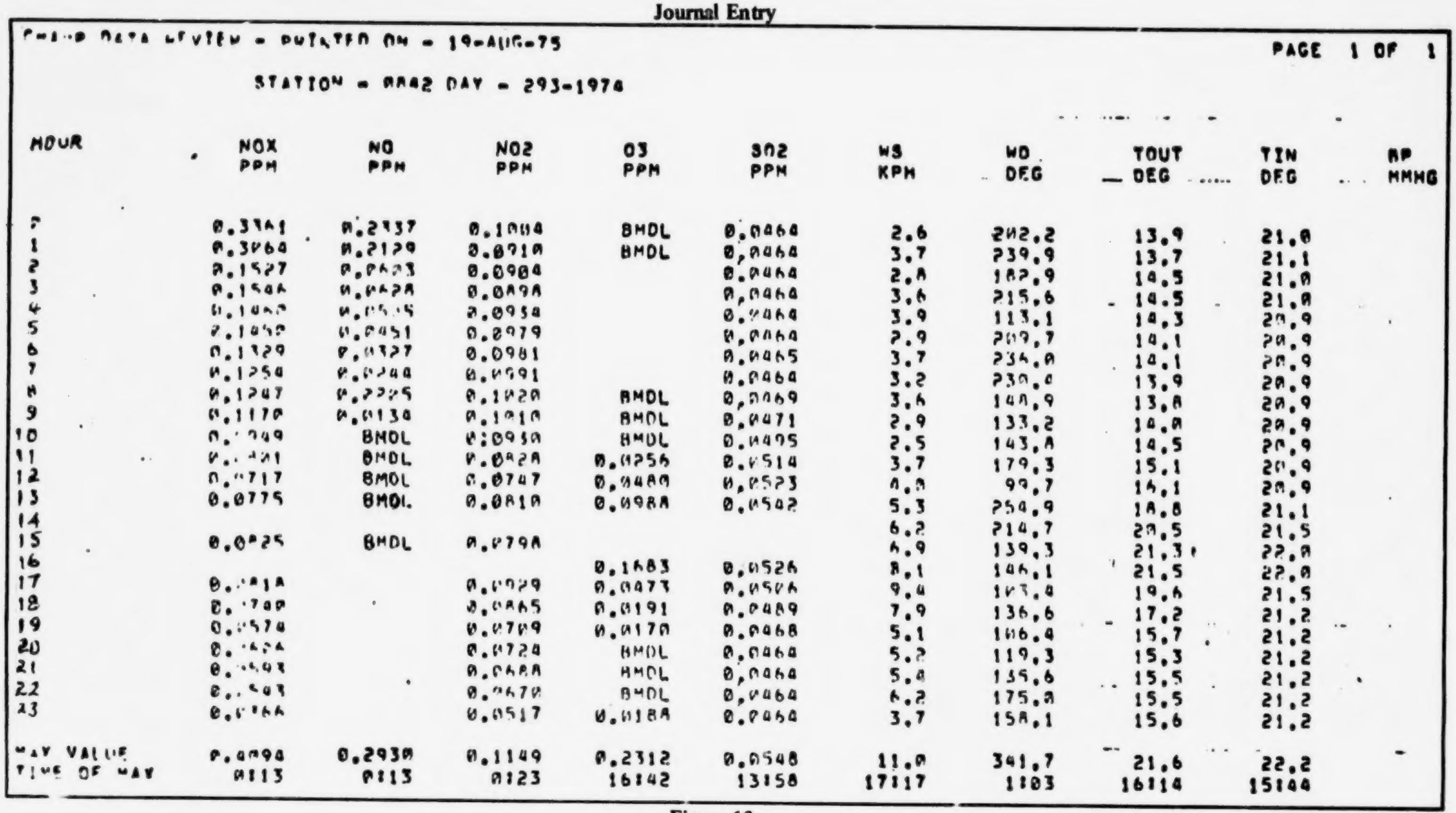


Figure 11
VALDAT-V



```
STATION -  0842 DAY 005-1975

                     JOURNAL ENTRIES
------------------------------------------------------------------

STATION -  842     DAY -    5              TIME - 14:56
WEST COVINA CHAMP 7682... MOANE OPER..,  A 1,2,3.. AEEE S 1,2,3..
03 MONITOR SPAN DRIFT OUT O TOLERANCE... ALL OTHER INSTRUMENTS.
LOOK GOOD ... STILL HAVING PROBLEMS WITH READING OUT CHL 39N
SO2 SAMPLE.. FLAME GOES OUT IN SO2 MONITOR WHEN C6 IS ACTUATED..
PULLED ROUTINE MAINT ON STA... MD2
```

Figure 12
Journal Entry



```
CHAMP DATA REVIEW - PRINTED ON - 19-AUG-75                                                        PAGE  1 OF  1

                 STATION - 0842 DAY - 293-1974

HOUR              NOX        NO         NO2        O3         SO2        WS       WD       TOUT      TIN      BP
                  PPM        PPM        PPM        PPM        PPM        KPH      DEG      DEG       DEG      MMHG

 0              0.3361     0.2337     0.1004      BMDL      0.0464      2.6     202.2     13.9     21.0
 1              0.3064     0.2129     0.0910      BMDL      0.0464      3.7     239.9     13.7     21.1
 2              0.1527     0.0623     0.0904                0.0464      2.8     182.9     14.5     21.0
 3              0.1506     0.0628     0.0898                0.0464      3.6     215.6     14.5     21.0
 4              0.1062     0.05[illegible]5     0.0954                0.0464      3.9     113.1     14.3     20.9
 5              0.1052     0.0451     0.0979                0.0464      2.9     209.7     14.1     20.9
 6              0.1329     0.0327     0.0981                0.0465      3.7     234.0     14.1     20.9
 7              0.1254     0.0244     0.0991                0.0464      3.2     230.0     13.9     20.9
 8              0.1247     0.2225     0.1020      BMDL      0.0469      3.6     140.9     13.8     20.9
 9              0.1170     0.0134     0.1010      BMDL      0.0471      2.9     133.2     14.0     20.9
10              0.[illegible]049       BMDL     0:0930      BMDL      0.0495      2.5     143.8     14.5     20.9
11              0.[illegible]81       BMDL     0.0828     0.0256     0.0514      3.7     179.3     15.1     20.9
12              0.0717       BMOL     0.0747     0.0480     0.0523      4.0      99.7     16.1     20.9
13              0.0775       BMOL     0.0810     0.0988     0.0542      5.3     254.9     18.8     21.1
14                                                                      6.2     214.7     20.5     21.5
15              0.0825       BMDL     0.0798                            6.9     139.3     21.3     22.0
16                                               0.1683     0.0526      8.1     146.1     21.5     22.0
17              0.[illegible]818                  0.0929     0.0473     0.0506      9.4     163.4     19.6     21.5
18              0.[illegible]700                  0.0865     0.0191     0.0489      7.9     136.6     17.2     21.2
19              0.0574                0.0709     0.0170     0.0468      5.1     106.4     15.7     21.2
20              0.[illegible]624                  0.0724      BMDL      0.0464      5.2     119.3     15.3     21.2
21              0.[illegible]493                  0.0688      BMDL      0.0464      5.4     135.6     15.5     21.2
22              0.[illegible]503                  0.0670      BMDL      0.0464      6.2     175.0     15.5     21.2
23              0.0766                0.0517     0.0188     0.0464      3.7     158.1     15.6     21.2

MAX VALUE       0.4094     0.2930     0.1149     0.2312     0.0548     11.0     341.7     21.6     22.2
TIME OF MAX      01:13      01:13      01:23      16:42      13:58     17:17     11:03     16:14    15:44
```

Figure 13
REVIEW

DEVELOPMENT OF THERMAL CONTOUR MAPPING

By George C. Allison

INTRODUCTION

The project to produce thermal contour maps was initiated in the spring of 1973 at the Environmental Monitoring and Support Laboratory-Las Vegas (formerly the National Environmental Research Center-Las Vegas). At that time, two separate capabilities existed for generation of computer contour plots and for collection of thermal data with an airborne infrared scanner. The intent of this project was to join these two capabilities into an automated system for generating thermal isopleths, or contour maps, as shown in Figure 1. The system would be applied to map thermal discharges into water bodies primarily for enforcement purposes.

The system diagramed in Figure 2 was conceived to meet the requirements for generating thermal contour maps. In order to complete the system, additional resources were required to supply the following capabilities:

- Analog recording capability aboard the aircraft.
- A ground station for analog to digital conversion.
- Software for data reduction and to interface with the contour plotting software.

While the system does not appear complicated, its development produced a number of interesting problems and alternatives.

AIRBORNE RECORDING

The requirement for analog recording aboard the aircraft was readily satisfied by purchasing a standard 14-channel analog recording unit. However, satisfactory recording of the scanner data was not immediately obtained. The first imagery produced from a recorded signal showed that a waviness had been introduced which was most noticeable at the trailing edge of the scan lines. The cause was found to be a very small oscillation of the tape recorder speed. Attempts to correct the problem through adjustments to the recorder were unsuccessful. Although the contour maps were never affected, the problem was minimized by recording at the maximum speed of 120 inches per second.

It had been hoped that a facility operated by a U.S. Energy Research and Development Agency contractor in Las Vegas could be used for digitizing the scanner data. However, that facility was used for digitizing ground motion data requiring a minimum interval of about .10 of a second, while the scanner signal was to be digitized at an interval of from 10 to 20 microseconds. Further investigation revealed the speed limitation of that system to be far short of that required, so our own digitizing facility was developed.

THE GROUND STATION

The alternatives considered for the ground station included both computer-based and nonprogramable digitizing facilities. The use of a minicomputer was chosen over a hard-wired facility primarily for flexibility and error detection capabilities. This decision led to another consideration: contract versus in-house development of the software. Due to the lack of available in-house personnel, an attempt was made to obtain both hardware and software from the computer manufacturer. The final result was a separate contract for software development with an individual recommended by the computer manufacturer.

When the minicomputer and software were ready for delivery, we were unable to accept the software because the "front-end" of the system consisting of a playback recorder and analog-to-digital converter was not yet available. In order to test the software for acceptance and to progress with the system, a simulator was developed in the place of the A-to-D converter. The simulator consisted of: (1) a square-wave generator, (2) a pulse generator to provide timing, and (3) a minimal amount of logic to permit data to be read.

The simulator proved to be more useful than anticipated. In addition to allowing for final checkout and acceptance of the digitizing software, it served as a continuously variable exerciser for finding the speed limitations of the ground station. It also made it possible to generate test tapes for development and checkout of the remainder of the system.

TEMPERATURE CONVERSION

The conversion from voltage to temperature is the most critical and potentially controversial phase of the system. In this area we have relied upon the technique developed by the NASA Earth Resources Laboratory in Mississippi.[1] This technique involves the use of temperature standards built into the scanner for determining the relationship between voltage and temperature, and the use of ground-truth data for determining atmospheric loss. The accuracy of the technique depends largely upon having constant atmospheric conditions over the spatial and temporal range of data collection.

Although the mathematics involved is quite simple, the temperature conversion software was meticulously prepared for both accuracy and efficiency. The concern for accuracy was not a concern for the computer accuracy, but for the programing accuracy. The system was developed with the ultimate purpose of providing information for enforcement action. If a case should be taken to court on the basis of information produced by this system, the data and software should be able to withstand detailed scrutiny by reviewing experts outside the U.S. Environmental Protection Agency.

Our concern for efficiency is necessary to keep computer time costs down. A typical area to be plotted might originally consist of 2 million or more digitized data points. The temperature conversion program uses several code loops in which instructions must be executed once per data point. Careless coding in these loops might double or triple the cost of the entire system.

CONTOUR PLOTTING INTERFACE

The contour plotting routine requires the input data to be in the form of an equally spaced orthogonal grid. Unfortunately, the scanner data contain an inherent geometric distortion that prevents direct input for contour plotting. Included with the contouring software is a routine to generate a suitable grid from irregularly spaced data; however, the routine is oriented toward the problems associated with creating a relatively dense grid from sparse input data. Since the scanner data are already denser than the grid required, most of the processing done by this routine would be unnecessary. The computer time used by the routine would be prohibitive unless most of the data available were discarded. This, however, would destroy the resolution and spatial accuracy of the map. All of these problems indicated that our own simplified grid generation program should be developed.

The grid generation program was developed with the same consideration for efficiency and accuracy as the temperature conversion program. The method used to generate the grid is to compute the coordinates of each grid point in the distorted coordinate system of the raw data. This locates the four nearest data points and the grid value is calculated by linear interpolation. The grid then can be used to produce a contour map centered about the flight line.

CONTOUR PLOTTING ROUTINE

The Surface Approximation and Contour Mapping (SACM) software is a general purpose set of programs originally developed for oil exploration applications. It has no features specifically developed for this application, nor does it seem to lack any features required. However, it has been the one part of the system where software failures have occurred. On occasions we have exceeded some of its limitations resulting in over-stored arrays and other failures. This situation occurs when large portions of the area contoured are land areas with many temperature variations. Since this software was obtained commercially and is massive in content, it is exceedingly difficult to identify and correct such problems. However, we have been successful at circumventing most problems through various options of the software. The most useful technique is the use of a routine to mask out polygons from the contour plot. This can be used to eliminate troublesome land areas from the plot.

CONCLUSION

This system has been in operation for a year and has required very few enhancements during that period. Four of the ten EPA regional offices are receiving thermal contour maps generated by this system. A comprehensive analysis currently being performed to assure the accuracy of the system has not yet indicated the need for any substantial software modifications. Most of the difficulty in producing a good contour map relates to the nature of the water body itself, and the conditions surrounding data collection. Generally there are sufficient options built into the system to overcome the difficulties.

REFERENCE

1 Boudreau, R.D., *Correcting Airborne Scanning Infrared Radiometer Measurements for Atmospheric Effects*, NASA Earth Resources Laboratory Report 029, Bay St. Louis, Mississippi 1972.

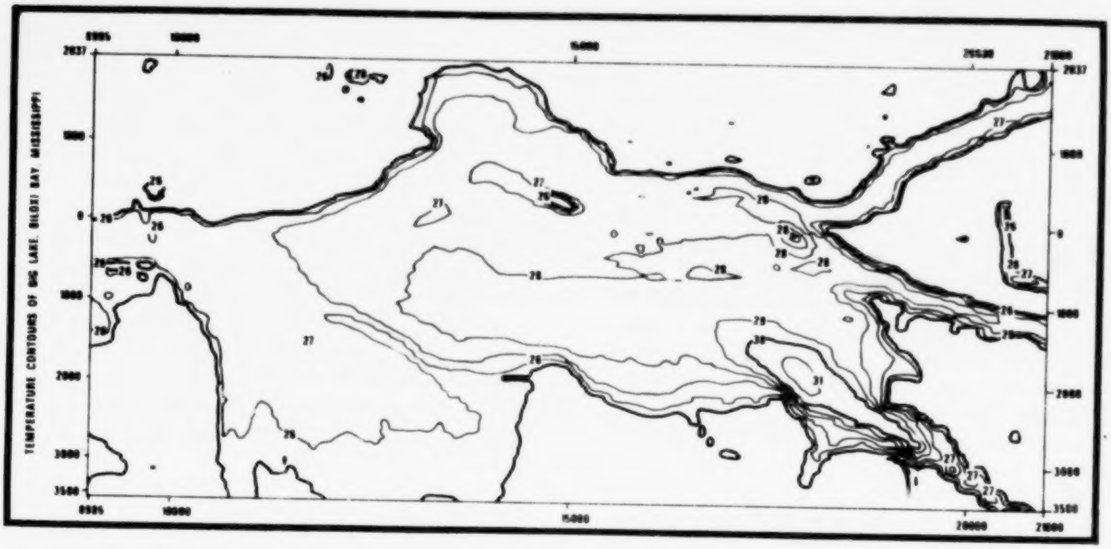

TEMPERATURE CONTOURS OF BIG LAKE, BILOXI BAY MISSISSIPPI
10000
15000
20000
21000
2037
1000
0
2000
3000
3500
27
28
29
30
31

Figure 1
Sample Thermal Isopleth

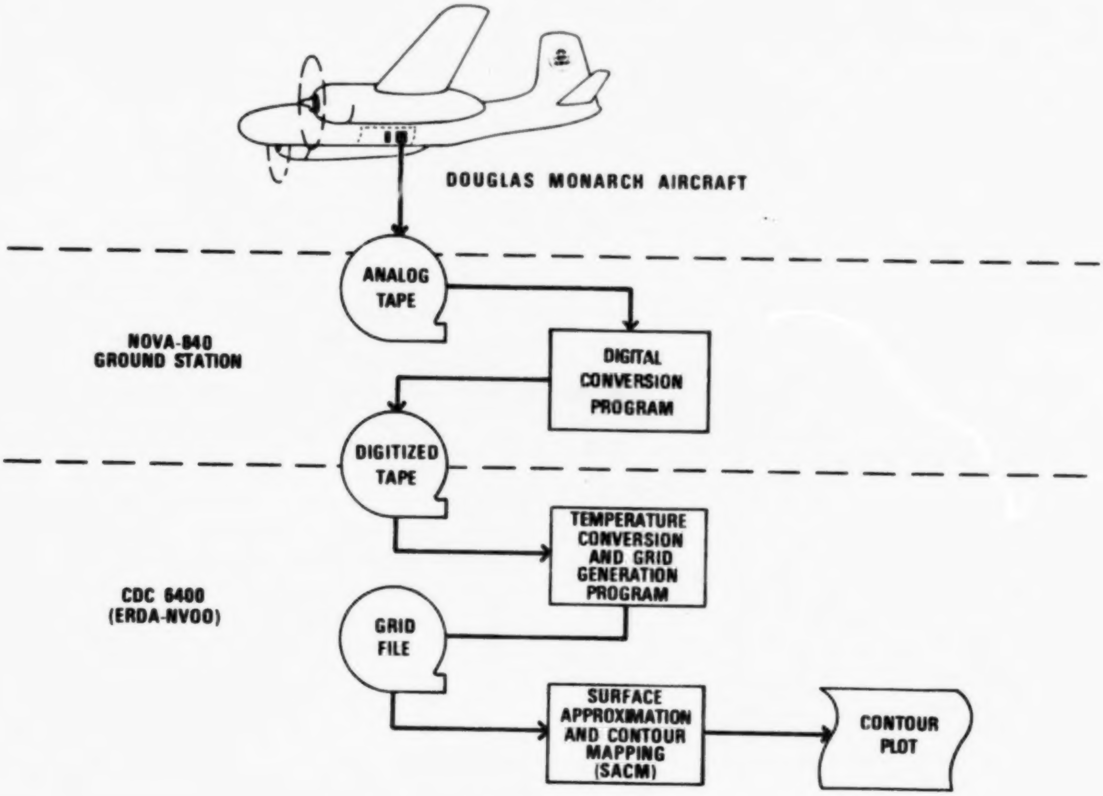

DOUGLAS MONARCH AIRCRAFT
ANALOG TAPE
NOVA-840 GROUND STATION
DIGITAL CONVERSION PROGRAM
DIGITIZED TAPE
TEMPERATURE CONVERSION AND GRID GENERATION PROGRAM
CDC 6400 (ERDA-NVOO)
GRID FILE
SURFACE APPROXIMATION AND CONTOUR MAPPING (SACM)
CONTOUR PLOT

Figure 2
System Flowchart

REMOTE SENSING PROJECTS IN THE REGIONAL AIR POLLUTION STUDY

By R. Jurgens

The St. Louis Regional Air Pollution Study (RAPS) was established in July 1972 with the objective of developing and evaluating mathematical air quality simulation models. Given the source emission data and meteorological conditions, these models would describe and predict the concentration, diffusion, and transport of pollutants over a regional area. One application of these models would be to assist State and local air pollution agencies in assessing the effectiveness of, and choosing between, alternate air pollution control strategies. Verified models could potentially reduce the requirements for actual pollution monitoring within a region.

A requirement of model evaluation is the availability of an extensive base of air quality and meteorological measurements and information on all processes that determine pollution concentration within the modeled area. A large research and development effort was established in St. Louis to meet this objective.

A number of experiments utilizing remote sensing instruments are part of the RAPS research effort. These include NOAA's acoustic sounder, EPA's Lidar systems, Lincoln Laboratories CO monitor, and the Regional Air Monitoring network. Descriptions of these remote sensing systems are included in this paper.

ACOUSTIC ECHO SOUNDER

An acoustic echo sounder was installed in the downtown St. Louis area early in 1975. The installed system is maintained by the Wave Propagation Laboratory of NOAA. The primary motivation for this project is a study of diurnal and seasonal changes in the urban boundary layer. Using acoustic radar thermal plumes, inversion layers, and their dynamic behavior have been studied. Recently, the Doppler frequency shift of scattered signals has been analyzed to determine wind velocities.

This remote sensing technique is based on the principle that acoustic waves propagating through the atmosphere are scattered by temperature fluctuations (variations in refractive index) and by fluctuations in the motion of the air.

The acoustic radar consists of three basic systems:[1] (1) a transmitting system that generates short, high-powered pulses of sound at a single frequency, (2) a receiving system that detects and amplifies the small fraction of incident pulse that is backscattered, and (3) an analyzing recording system—usually a facsimile recorder which produces time-height profiles of the echoes and perhaps a multichannel analog magnetic tape recorder. Typical sounder parameters in use in St. Louis are: transmitted carrier frequency 2950 Hz, transmitted acoustic power 10 W, transmitted pulse duration 200 ms, pulse interval 5 sec, and receiver bandwidth 30 Hz.

Currently, only the facsimile-record echo data are being analyzed. Analysis is based on pattern recognition techniques developed by Clark and Bendun.[2] Thirteen general classification patterns have been defined, and these are used with slight modification for each new site. An example of a continuous pattern categorization of acoustic sounder facsimile records is shown in Figure 1. From data like these, it is possible to study the diurnal trends and frequencies of occurrences of the various patterns. With the aid of supporting ground level and vertical profiles of meteorological variables, it is hoped to fully describe the prevailing atmospheric condition causing the various pattern types.

LIDAR

Both ground and airborne Lidar[3] (light detection and ranging) studies have been conducted in St. Louis during the 1974 and 1975 summer field intensives. These Lidar systems augment experimental studies of the boundary layer structure by determining mixing layer heights over the urban area, especially during the morning and evening transition periods which are characterized by discontinuous and/or fluctuating changes in the mixing height. The airborne system also has been used in determining the dimensions of plumes. The airborne Lidar has the unique capability of being able to make many measurements over large geographic areas in a relatively short period of time.

Both Lidar and long-path CO monitoring systems use lasers (light amplification by stimulted emission of radiation) for their source of pulses.[4] Whereas the principle of operation of the CO laser is based on molecular resonant absorption, the Lidar system, often referred to as laser radar, measures the backscattered energy of a pulse transmitted through the lower atmosphere. The pulse is scattered off aerosols or off dispersed solid or liquid particles. The principle of pulse generation is the same for both laser systems. A contained system of atoms is "pumped" to an active or excited stage. Laser action is initiated when excited atoms spontaneously decay to lower energy levels, emitting photons in the process. Photons trapped within the container trigger other emissions which are in phase with the triggering photons. The cascaded emissions are contained within the material long enough to produce the laser beam.

The optical system of the ground-based Lidar is shown in Figure 2. The transmitter consists of a Q-switched air-cooled, pulsed ruby with wavelength 6943A (deep red). Since the angular resolution of the Lidar is determined by the transitted beam divergence, 6-inch diameter (38-cm Fresnel lens on the airborne Lidar) collimating optics are used to reduce the laser beam divergency and to produce an output beamwidth of 35 mrad. The corresponding spatial resolution of this beam is 0.5 at a range of 1 km in the crossbeam direction, and about 2.3 m in range. The maximum firing rate, limited by the cooling rate is 12 pulses per minute. In the receiver, a multilayered narrow-band filter is inserted to reduce the output noise level produced by solar radiation scattered into the receiver field of view. During operation, a compressed air-driven turbine rotates the laser Q-switch prism at 500 r/s. Upon receipt of a fire signal, a synchronizing generator triggers the flash lamp in step with a signal from the rotating prism. A capacitor bank charged to 3 kv supplies energy for the laser flash lamps.

Detected signals from both Lidar systems are output onto strip charts and also passed through A/D converters for storage on magnetic tape. The strip chart data are subsequently digitized for storage on magnetic tape. Analysis and plotting are done on a large batch computer.

An example of airborne Lidar data from an industrial plant is shown in Figure 3. The Lidar returns are from the north to south transverse over the Union Electric Sioux powerplant. With a northeast wind, there were little or no aerosols upwind of the plant.

LONG-PATH LASER MONITORING OF CO

During the 1974 and 1975 summer intensive field experiments in St. Louis, a tuneable semiconductor diode laser system mounted in a mobile van was used to make long-path (0.3-1 km) integrated measurements of CO. The system was developed by MIT Lincoln Laboratories with funding from EPA and NSF.[5]

The basis for the operation of this system is the absorption of the laser radiation by gas molecules.[6] The measured intensity of the laser beam at the detector can be related to the integrated concentration of the target gas over the path length. The essential components of the laser system are shown in Figure 4. The diode laser is mounted in a closed-cycle cryogenic cooler which is maintained between 10-20 K. The laser emission is collimated by an aluminum-coated parabolic mirror, 12 cm in diameter. The beam is transmitted down range to a remote retroreflector which reflects it back to the parabolic mirror and then onto an infrared detector situated behind a calibration cell. A sophisticated spectroscopic technique was devised to minimize the effects of atmospheric turbulence on system sensitivity. Detector output is recorded on a strip chart for subsequent digitizing on a Hewlett Packard 9864A and storage on cassette tapes. Analysis and plotting are on Hewlett Packard equipment at Lincoln Laboratories.

The laser source used is one of the Pb-salt types. It was tailored chemically to operate in the 4.7-μm wavelength region (infrared) in close coincidence with the fundamental vibrational band of CO entered at 2,145 cm^{-1}. Exact frequency matching and tuning through absorption lines is achieved by varying the injection current which changes the junction temperature, and thus the laser wavelength.

The St. Louis experiments were the first field measurements of this newly developed technology. Besides demonstrating this technology, the RAPS long-path CO laser experiment is being used to study pollutant variability around selected Regional Air Monitoring (RAMS) sites. The laser data are also being compared directly with RAMS data. An example of correlation between RAMS site 108 data and the laser monitor is shown in Figure 5. The two large increases in CO at about 7:30 and 8:30 a.m. represent plume crossings from a slag-processing plant in Granite City. Wind direction strongly affects the correlation between the two experiments and must be considered when comparing data.

The monitoring capability of the diode laser is being expanded to include NO, O_3, NH_3 and perhaps to weakly absorbing SO_2.

REGIONAL AIR MONITORING SYSTEM (RAMS)

RAMS[7] is the ground-based air pollution, meteorological, and solar radiation measurement network of RAPS. It consists of 25 stations situated in and about St. Louis. Figure 6 shows the placement of the 25 stations and the telephone trunk lines which connect them to the central facility at Creve Coeur (CCF in figure). RAMS is a sister system to CHAMP described in a paper by Marvin Hertz at this workshop.

Although RAMS is not a remote sensing project in a strict sense, the design philosophy allowed for untended operation of the remote sites except for routine maintenance. Features incorporated into the remote site to implement this philosophy include:

- Automatic power fail and automatic restart
- Backup storage for up to 3 days on magnetic tape
- Software digital commands used to remotely control the calibration of the gas analyzers
- 77 status sense bits which monitor system performance and associated support characteristics.

Operation of these features has proven successful in reducing the frequency of required maintenance and the manpower requirements needed to operate the stations.

Maintenance of telecommunications between the central computer facility and the remote sites is required for automatic operation of the system. The actual communication is through Novation modems running with Bell 202 type compatibility. Communication rates are 1200 baud in both directions using ASCII character formats. In addition to the parity function provided by ASCII, each transmission by the remote or central sites includes a check sum for greater redundancy in error detection. Bit error rates are on the order of 2 in 10**7 except during periods of electrical storms. Experience with the RAMS system indicated that between 5 to 10 percent of potential data is lost because of telecommunication problems.

ACKNOWLEDGMENTS

EPA investigators and contacts for the projects are:

- Acoustic sounder:
  Frank A. Schiermeier
  Regional Air Pollution Study
  11640 Administration Drive
  Creve Coeur, Missouri 63141

- Lidar:
  James L. McElroy or J.A. Eckert
  Environmental Monitoring
  & Support Laboratory
  Environmental Protection Agency
  Las Vegas, Nevada 89114

- CO monitor:
  William A. McClenny
  Environmental Research Science Laboratory
  Environmental Protection Agency
  Research Triangle Park, North Carolina 27711

- RAMS
  James A. Reagan
  Regional Air Pollution Study
  11640 Administration Drive
  Creve Coeur, Missouri 63141

REFERENCES

Acoustic Sounder

1 Mandics, P.A., Hall, F.F. and Owens, E.J., *Observations of the Tropical Marine Atmosphere Using an Acoustic Echo Sounder During Gate,* AMS, 16th Radar Meteorology Con., 1975.

2 Clark, G.H. and Bendun, E.O.K., *Meteorological Research Stuides at Jervis Bay, Australia.* Australian Atomic Energy Commission Report AAEC/E309; ISBN 064299B423; July 1974.

Lidar

3 Eckert, J.A., McElroy, J.L., Bundy, D.H., Guagliardo, J.L., and Melfi, S.H., *Downlooking Airborne Lidar Studies*, August 1974 (to be published as an EPA report).

4 Johnson, W.B., Jr. and Uthe, E.E., *Lidar Study of Stack Plumes*, SRI, June 1969.

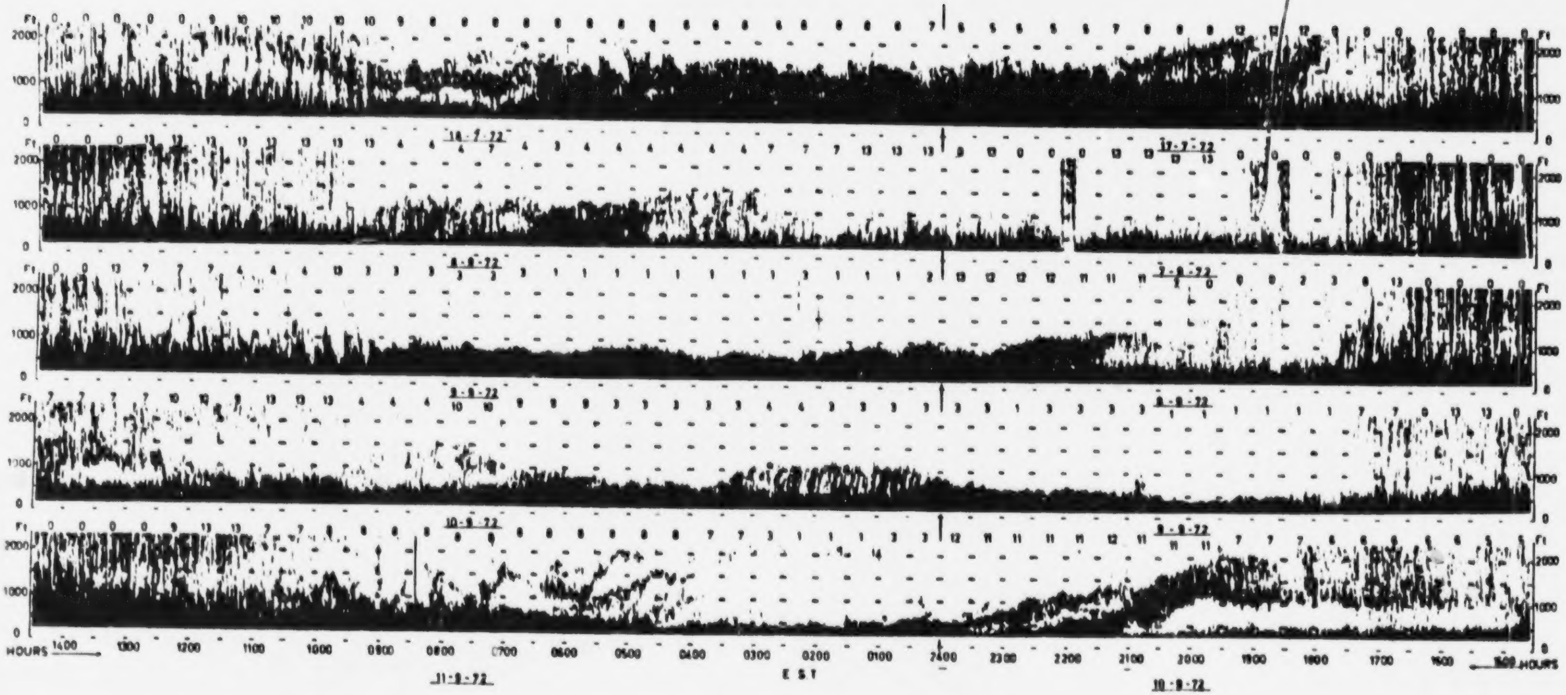
CO Monitor

5 McClenny, W.A. *Ambient Air Monitoring Using Long-Path Techniques*, Paper 28-6, International Conference on Environmental Sensing and Assessment, September 1975 (to be published).

6 Hinkley, E.P. *Long-Path Ambient Air Monitoring with Tuneable Lasers in St. Louis,* Lincoln Laboratories, MIT, January 1975.

RAMS

7 Myers, R.L. and Reagan, J.A. *Regional Air Monitoring System at St. Louis, Missouri,* International Conference on Environmental Sensing and Assessment, September 1975 (to be published).



Note: As marked, not all the categorizations are correct.

Figure 1
An Example of a Continuous Half-Hourly Pattern
Categorisation of Monostatic Acoustic Sounder
Facsimile Records Taken over Several Days

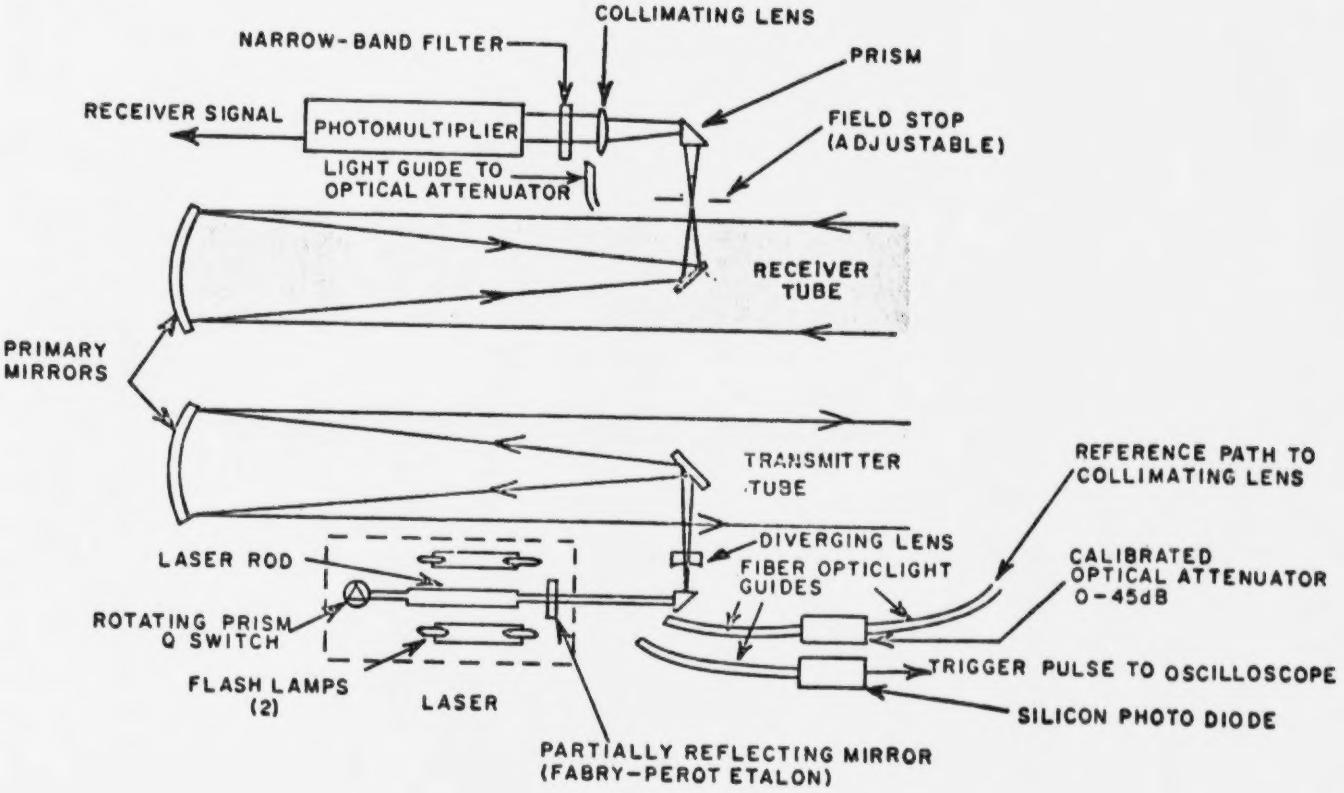
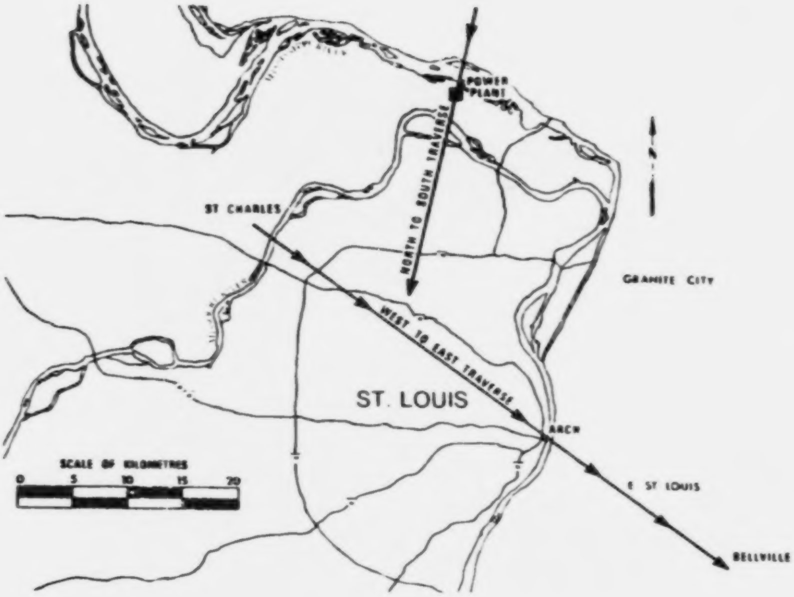

COLLIMATING LENS
NARROW-BAND FILTER
PRISM
RECEIVER SIGNAL
PHOTOMULTIPLIER
FIELD STOP (ADJUSTABLE)
LIGHT GUIDE TO OPTICAL ATTENUATOR
RECEIVER TUBE
PRIMARY MIRRORS
TRANSMITTER TUBE
REFERENCE PATH TO COLLIMATING LENS
DIVERGING LENS
LASER ROD
FIBER OPTICLIGHT GUIDES
CALIBRATED OPTICAL ATTENUATOR 0-45dB
ROTATING PRISM Q SWITCH
TRIGGER PULSE TO OSCILLOSCOPE
FLASH LAMPS (2)
LASER
SILICON PHOTO DIODE
PARTIALLY REFLECTING MIRROR (FABRY-PEROT ETALON)

Figure 2
Optical System for Ground-Based LIDAR


POWER PLANT
NORTH TO SOUTH TRAVERSE
ST CHARLES
WEST TO EAST TRAVERSE
GRANITE CITY
ST. LOUIS
ARCH
E ST LOUIS
BELLVILLE
SCALE OF KILOMETRES
0
5
10
15
20
N

Map of St. Louis Showing Lidar Traverses on August 19-20, 1974

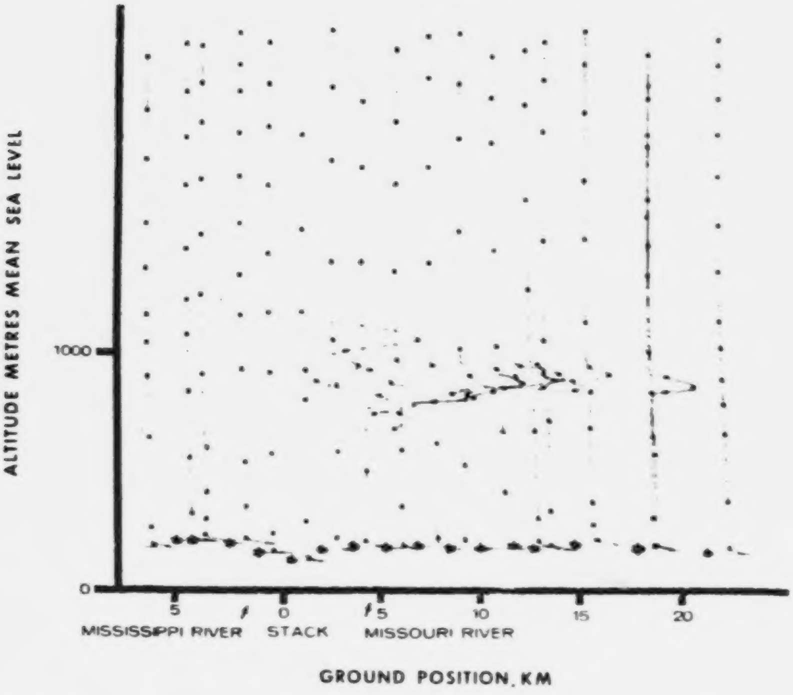


Figure 3
LIDAR Return Signals from North to South Traverse
Over Power Plan: North of St. Louis, Mo.

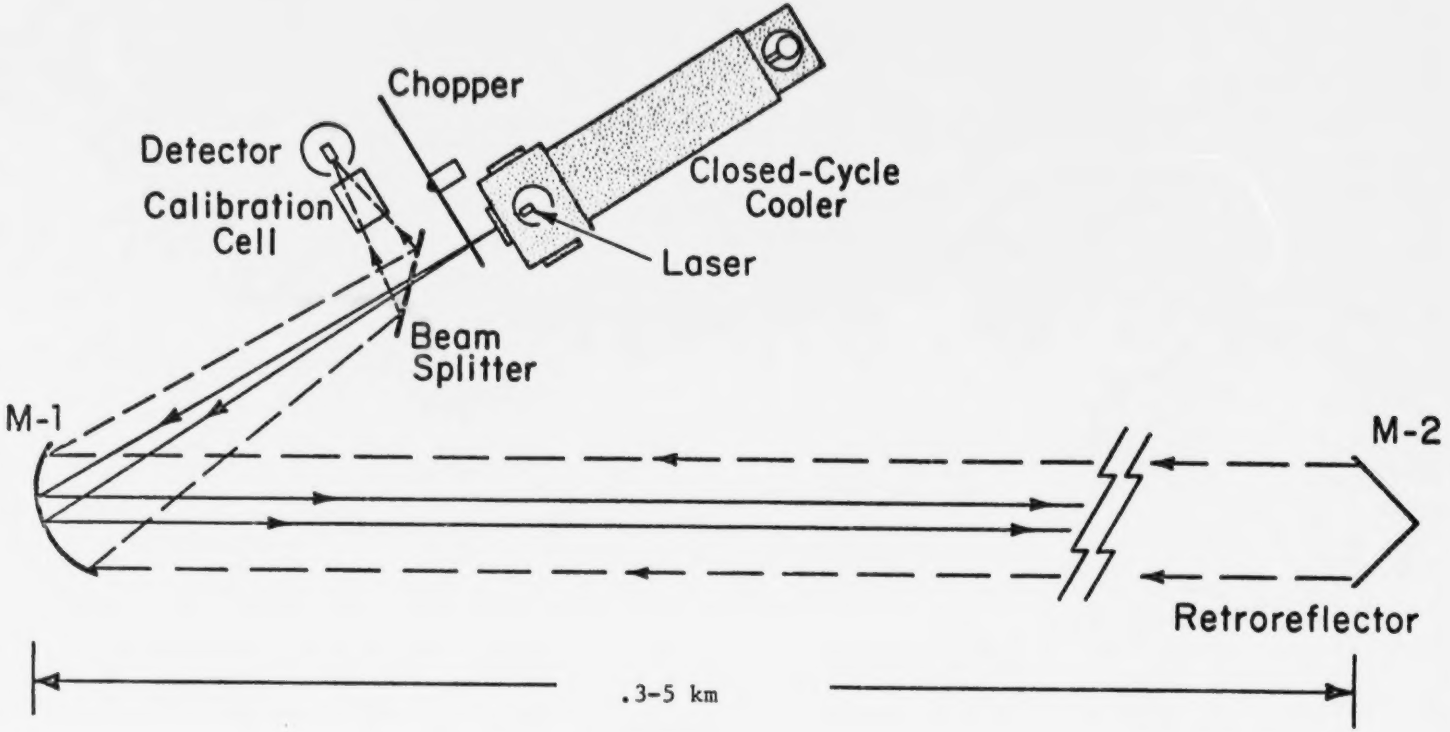

Chopper
Detector
Calibration Cell
Closed-Cycle Cooler
Laser
Beam Splitter
M-1
M-2
Retroreflector
.3-5 km

Figure 4
Optical System for Laser Monitoring

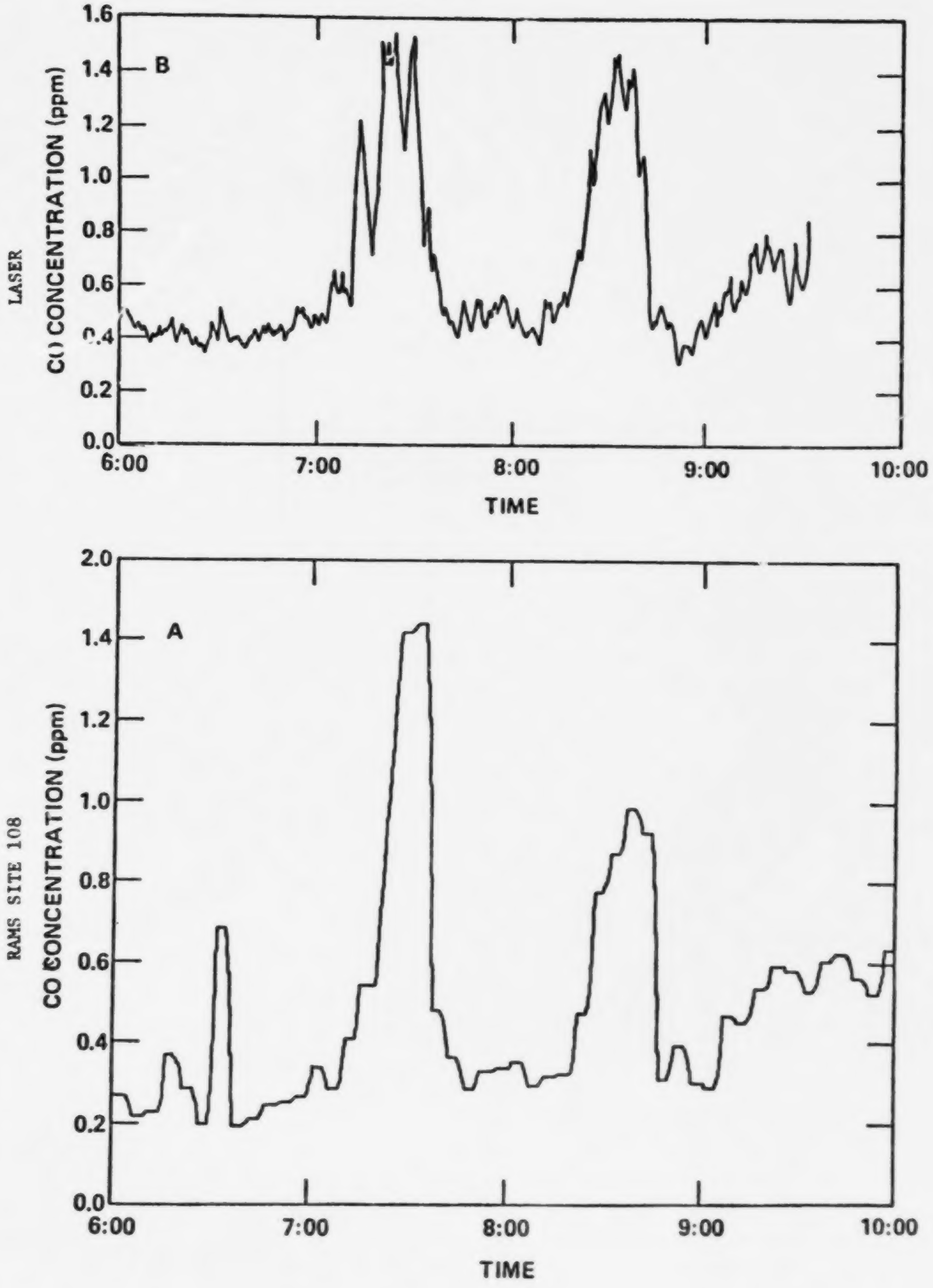


Figure 5
Correlation of LASER and RAMS
CO Measurements

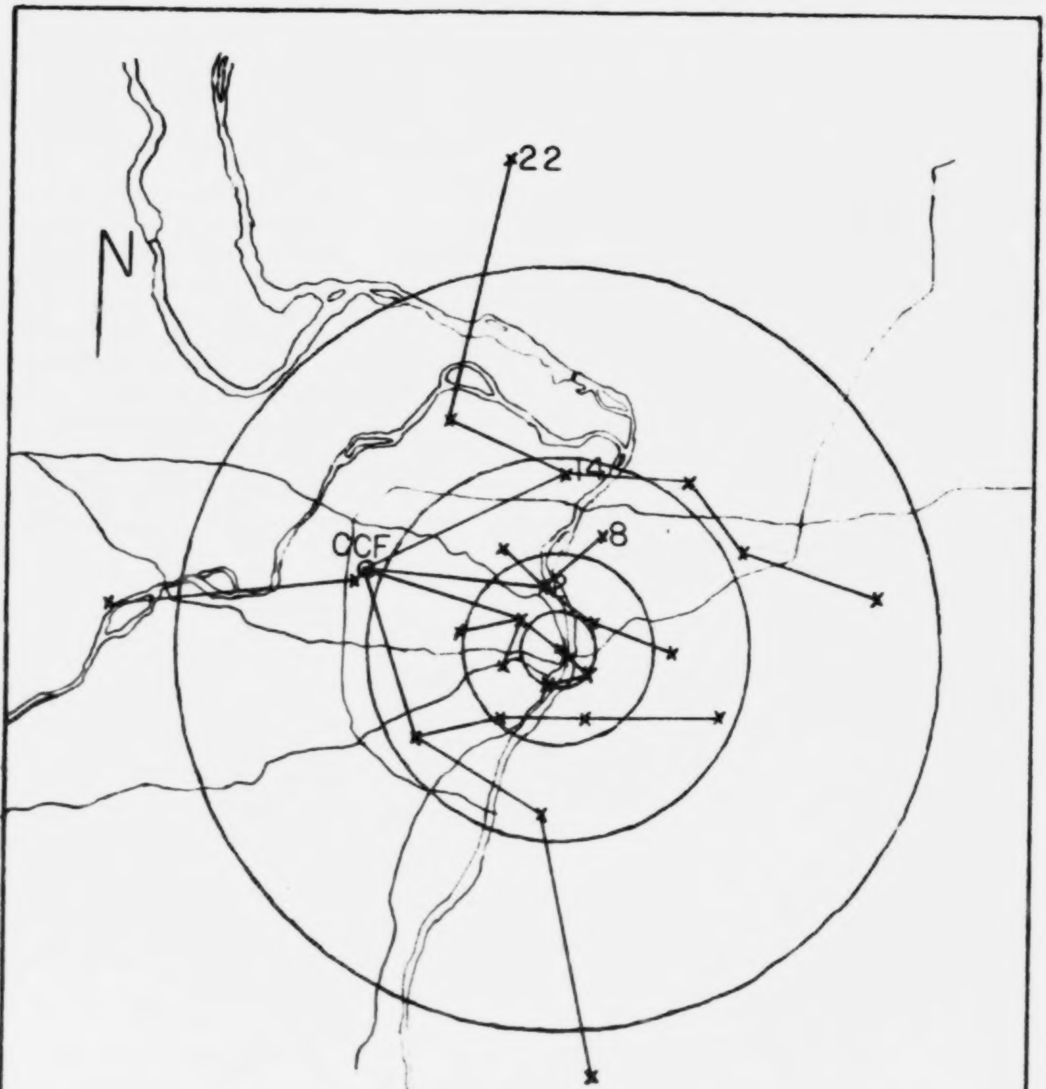


Figure 6
25 RAMS Stations With Trunk Lines to the Central Facility at Creve Coeur (CCF)

AUTOMATIC DATA PROCESSING REQUIREMENTS IN REMOTE MONITORING

By J. Koutsandreas

INTRODUCTION

The application of remote monitoring technology to environmental monitoring has taken a quantum jump since the advent of the space age; we have seen the simultaneous development of advanced sensor systems and platforms to carry them. This remote monitoring technology encompasses imagery and other forms of data acquired by a wide assortment of sensors aboard aircraft or orbiting spacecraft, or automated data collection systems acquired through telemetry links. Information processing techniques range from conventional visual interpretation to sophisticated computer interactive (man-machine) systems. These techniques have provided a means of reducing information to useful formats, including base-map overlays, electronically displayed color-coded thematic maps or, for some applications, computer-generated maps and tabular data.

In this paper, some of the useful techniques of remote environmental monitoring are discussed, with an emphasis on automatic data processing. Included is a brief description of the sensor systems which gather the data, and a discussion of some of the data requirements and a few applications.

SENSORS

When carried aboard aircraft or spacecraft, remote sensors offer a synoptic overview which is not achieved by ground survey methods. Observations of the total scene are recorded as an image, and present a visual set of data patterns, not merely the group of data points which would have been collected by ground methods. The remote sensor sampling technique is an unobtrusive way of gathering data. The mere presence of a ground survey team, for example, investigating potential sites for development may result in the spread of unfounded rumors and may cause unwarranted adverse reactions which would hinder further site evaluation. The final images of remote sensors have a very high information density compared with graphic, textual, or electronic storage media. Thus, the remote sensor presents a more comprehensive picture of the area than conventional field methods. Although the level of detail recorded by a sensor may not be as great for a small area as ground observations would be, the sensor record affords a valuable overview in a manageable form. The cost/benefit ratio between overhead coverage and ground traverses for a given area greatly favors the remote sensing approach, except in cases where investigation of only a very small area is required. However, the remote sensors can also indicate where to concentrate more detailed in situ sensing and sampling.

The remote monitoring systems that provide and will continue to provide this Agency with environmental data are listed below, with their respective data storage medium:

Sensor Type	Data Storage
CAMERAS	Film
MULTISPECTRAL SCANNERS	Tape/Film
SPECTROMETER/RADIOMETER	Tape/Film
LASER/LIDAR	Tape
AUTOMATED IN SITU SENSORS	Tape

Photographic systems are still the most important of all the remote sensors. These systems include metric cameras and panoramic cameras, which are being used for conventional monitoring by DOD, NASA, EPA, and other Federal agencies. Routine requirements for the processing of camera films are not included in this paper.

Multispectral Scanner

A multispectral scanner (MSS), Figure 1, is a device which provides data in a multiband mode similar to that obtained from multiband camera systems. The major operational difference is that a multispectral scanner is an electro-optical instrument. The scanners are configured with single or arrayed detectors which sense the incoming image from a collector optic (scanning mirror). Scanners look at a single "spot" of the area at any given instant in time. The spot is scanned laterally to produce a line of imagery, also shown in Figure 1. The forward motion of the aircraft or spacecraft **collects** successive lines to produce a swath of the scene. These lines are translated into imagery. The incoming optical signal image is converted to a modulated electrical signal which can be either recorded on tape for later reproduction or used to vary a point source of illumination to photographically record the image on film.

In the case of multiple-detector arrays, each detector or group of detectors is designed to provide an optimum response over a discrete portion of the electromagnetic spectrum. In this way, a single overflight with a multispectral scanner can provide a number of simultaneous "filtered" recordings, from which a number of different spectral terrain characteristics may be detected.

The scanner technique is always applied to thermal infrared (IR) sensing since no film emulsion is capable of direct thermal IR recording. Thermal surveys, used to record either relative heat contrasts of surface objects or absolute thermal values (with ground control calibration inputs), have been extensively employed for application in studies such as those of thermal pollution and near surface geologic structures as seen in Figure 2. Attempts are being made to devise workable techniques for monitoring oil spills with thermal IR scanners. Multispectral scanners can detect 256 levels of gray, as compared with camera systems which can record only 30 shades of gray. This is a great advantage when looking for very subtle changes in water or land; imagery is presented with a greater spectral response than what is obtainable from camera films.

Spectrometer/Radiometer

The spectrometer and radiometer are passive devices which can measure spectral radiances over wavelengths of 0.3–14 μm and record the radiance levels. The outputs are line plots on a coordinate system. The spectral resolution of the spectrometer is much narrower than the radiometer and can detect individual pollutants such as SO_2 and O_3 by looking at the absorption spectra of the pollutants at specific wavelengths.

LASER/LIDAR

The LASER converts input pump power into coherent optical out power. The output is a coherent radiation which can relate a variety of environmental factors because of the following LASER characteristics:

- Narrow frequency
- Highly directional
- High intensity
- Constant phase.

A LIDAR profilometer has been developed and is shown in Figure 3. The terrain profile of strip mine areas reveals quantitative information such as elevation, slope, tailings, and revegetation characteristics. Another LIDAR application was demonstrated in the St. Louis RAPS program. An isoscattering contour plot over St. Louis, Missouri, made by flying a LIDAR on a helicopter, is shown in Figure 4. Notice the high density of particulates east and west of the river. Contour plots of this type, made within an hour, are not cost effective using in situ methods.

Automated In Situ Sensor Systems

These systems depend on electronic relays to transmit information to a central location for storage and analysis. This system will provide the capability of collecting vast streams of data automatically from in situ sensors (e.g., pH, D.O., heavy metals, etc.) located at remote or inaccessible locations on the surface. The value of such a system is that it facilitates the recovery of continuous data in regions which have, until now, required extensive field surveys to acquire even a few data points. Each sensor/transmitter is designed to continuously sample and record data and, by receipt of a coded signal or on a prescribed time schedule, transmit these values to a satellite, an aircraft, or a ground station within the line-of-sight of the in situ sensor mounted on a platform in water or on land.

The data collection system promises to record the continuous data required for environmental baseline studies and will make possible early warning of such incidents as floods, earthquakes, forest fires, oil spills, and offshore dumping violations.

DATA PROCESSING AND PREPARATION

In order to extract all vital information, remotely monitored data requires a complete processing capability. This usually includes computer operations, systems analysis, and applications programing for problem definition and mathematical analysis. The data reduction system consists of computers with a complete set of standard peripheral devices, special devices for image digitizing, and display. Multispectral processing of the data requires an optimum hardware/software configuration, accurate algorithms, and data processing techniques. Data preparation includes photographic and processed electronic data collection, preparation, and documentation.

REMIDS

In the Environmental Monitoring and Support Laboratory in Las Vegas, Nevada, the Remote Micro Imagery Data System (REMIDS) will provide an efficient method for storage and retrieval of interpreted remote sensing imagery in high resolution microform. This system, shown in Figure 5, is oriented toward supporting various existing and anticipated EPA regulatory permit programs which require periodic field inspections. A central index of the stored data is maintained which can be accessed via remote terminal devices.

The proposed system is currently composed of three programs which have been designed to allow maximum flexibility of output. A fourth program will be added once the system has been finalized. This fourth program will be an edit update package and will be used for file maintenance purposes.

The following is a brief description of the three existing programs:

1. Aperture Card: This program prints the aperture card and creates a master file.

2. Selection Program: This program allows the user to selectively query the master file and to determine what information is available, primarily in a specific geographic area. The user can select in any of the following fields:

- State name
- County name
- City name
- Facility name
- Receiving waters
- Standard Industrial Codes (SIC) - This selection can be on the first, second, third, or all four digits.

Currently, the selective program is set up to allow up to 20 different names in each field. The number is arbitrary and can be expanded easily. The selection process is mutually exclusive. For a record to be selected, it must meet all selection criteria. For example, to select all the facilities in Pennsylvania, the user would use one selection specifying Pennsylvania. To get all records in Washington County, Pennsylvania, the user would specify Pennsylvania/Washington. If Pennsylvania were left off, the user would get the records for Washington County in any State that has a county of that name. The selected records can be printed in the standard sequence, which is Facility Name within State Name, or in any sequence desired by the user (County, SIC, Major Industry Code, Receiving Water, or Discharge Number).

3. Polygon Selection: This program selects all records which fall within a polygon specified by up to 20 latitude/longitude points. The report options for sequence and format are the same as the selection program.

When the proposed system is finalized, cookbook instructions will be provided to the EPA Regions.

COMPUTER IMAGE PROCESSING

Modern technology utilizes all types of pictures, or images, as sources of information for interpretation and analysis. These may be portions of the earth's surface viewed from an aircraft or an orbiting satellite. The proliferation of these pictorial data has created the need for a vision-based automation that can rapidly, accurately, and cost effectively extract the useful information contained in images. These requirements are being met through the new technology of image processing. A typical system is shown in Figure 6.

Image processing combines computer applications with modern image scanning techniques to perform various forms of image enhancement, distortion correction, pattern recognition, and object measurement. This technology overcomes many of the inherent difficulties associated with the human analysis of images or objects; however, it is based upon the same fundamental principles as visual recognition in human beings. Although the actual visual process is physiologically complex, the basic mechanism of vision uses the eyes and brain as an automatic information interpreting system. The eyes receive stimuli in the form of visual light, and the brain processes and interprets this input for the observer of the image. The human visual system can be simulated using an electronic scanner for its eyes, similar to a television camera, and a high-speed digital computer for its brain. This type of system can "see" images through the scanner and, by means of the programed capabilities of the computer, can manipulate the images in various ways that contribute to extracting desired information which is usually not apparent to the untrained observer.

Through various aircraft and satellite programs, a profusion of remotely sensed images are constantly being acquired for use in the monitoring of pollution. Image processing technology is providing the ability to rapidly and cost effectively extract the abundance of useful information embodied in these remotely sensed data.

In conclusion, I have explained the necessity of ADP in remote sensor data processing. It is only through the use of computer technology that EPA scientists will be able to fully exploit the outputs of remote sensors and realize their potential in the area of environmental monitoring.

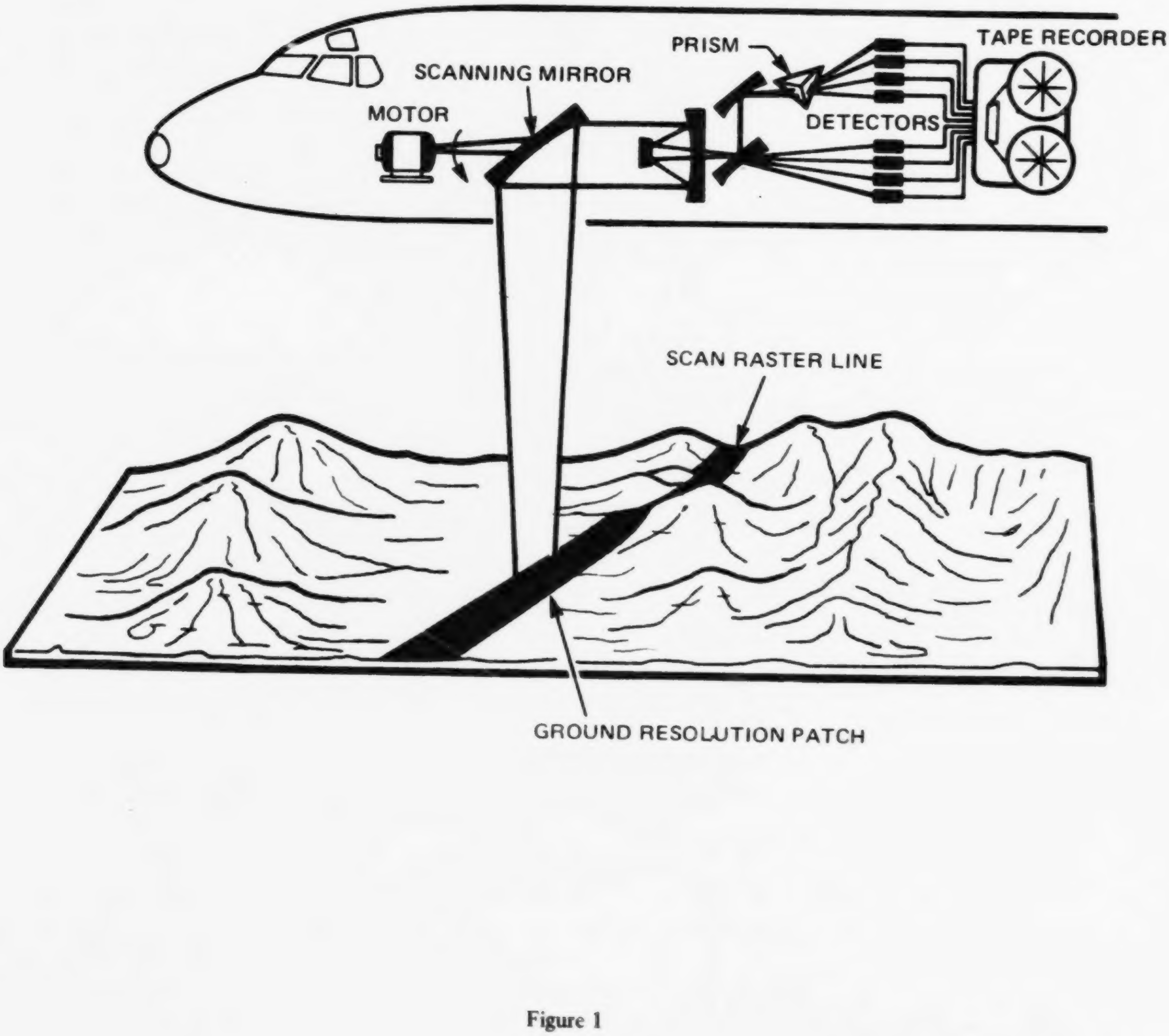
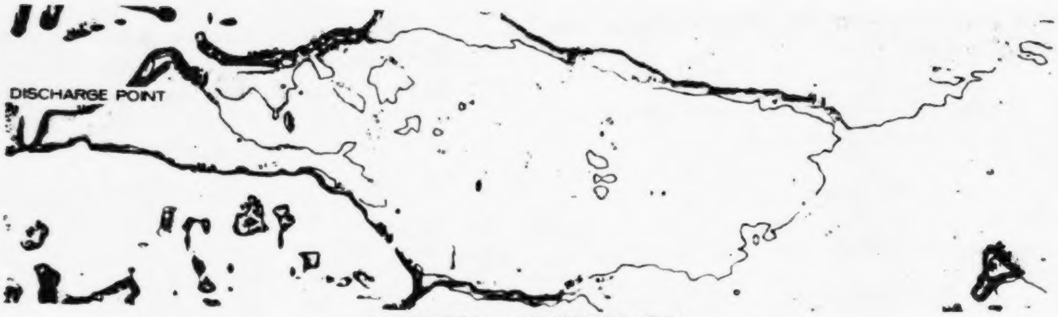

PRISM
TAPE RECORDER
SCANNING MIRROR
MOTOR
DETECTORS
SCAN RASTER LINE
GROUND RESOLUTION PATCH

Figure 1


DISCHARGE POINT

ISOTHERMAL CONTOURS (°F)

13:31:30 EDT
13:32:00 EDT
DISCHARGE POINT

THERMAL INFRARED IMAGERY

DISCHARGE POINT
LEE RIVER
MOUNT HOPE BAY

COLOR AERIAL PHOTOGRAPHS

0
1000
2000
FEET (approximately)

NERC-LV PROJECT 7502
FLOWN SEPTEMBER 1, 1974

NEW ENGLAND POWER SERVICE CO.
BRAYTON POINT POWER STATION
SOMERSET, MASSACHUSETTS

NERC-LAS VEGAS

Figure 2

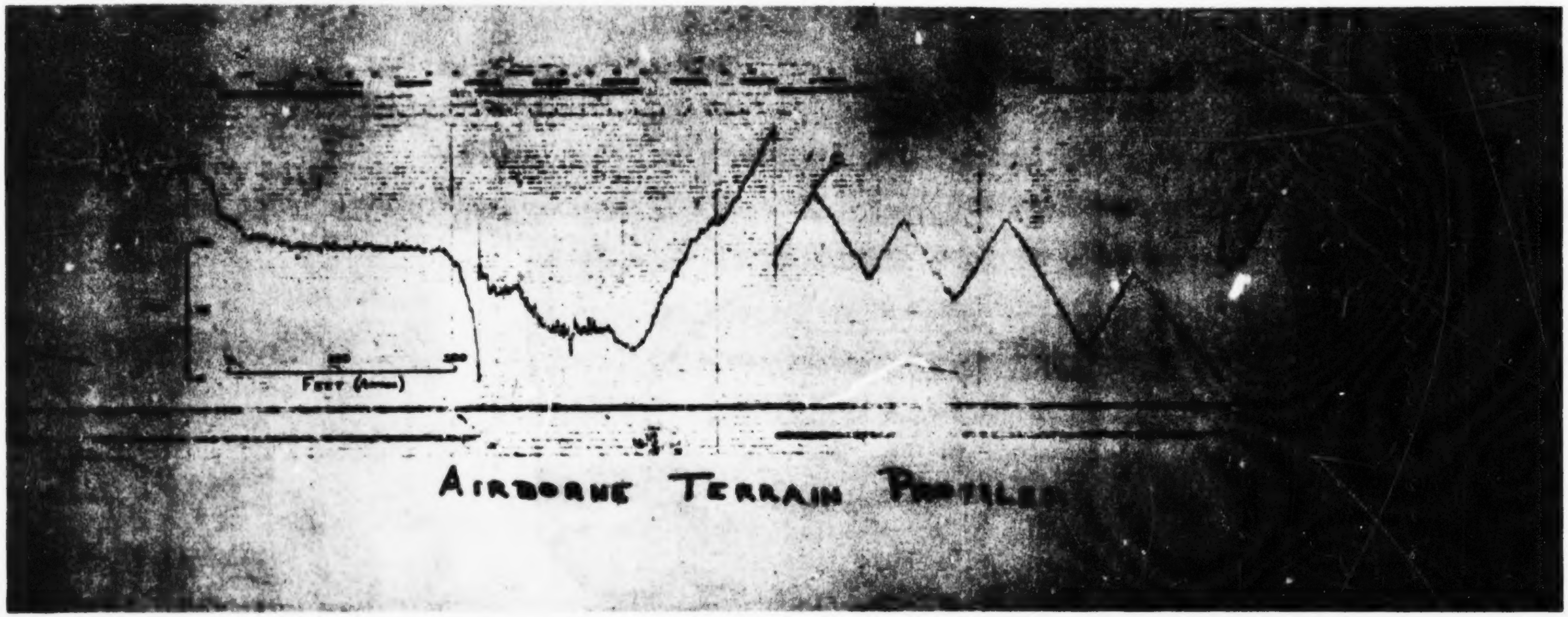


Figure 3

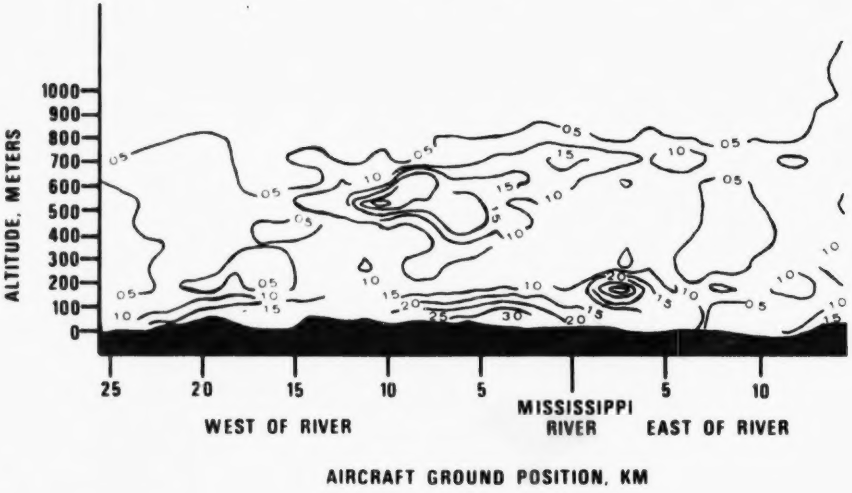


Figure 4

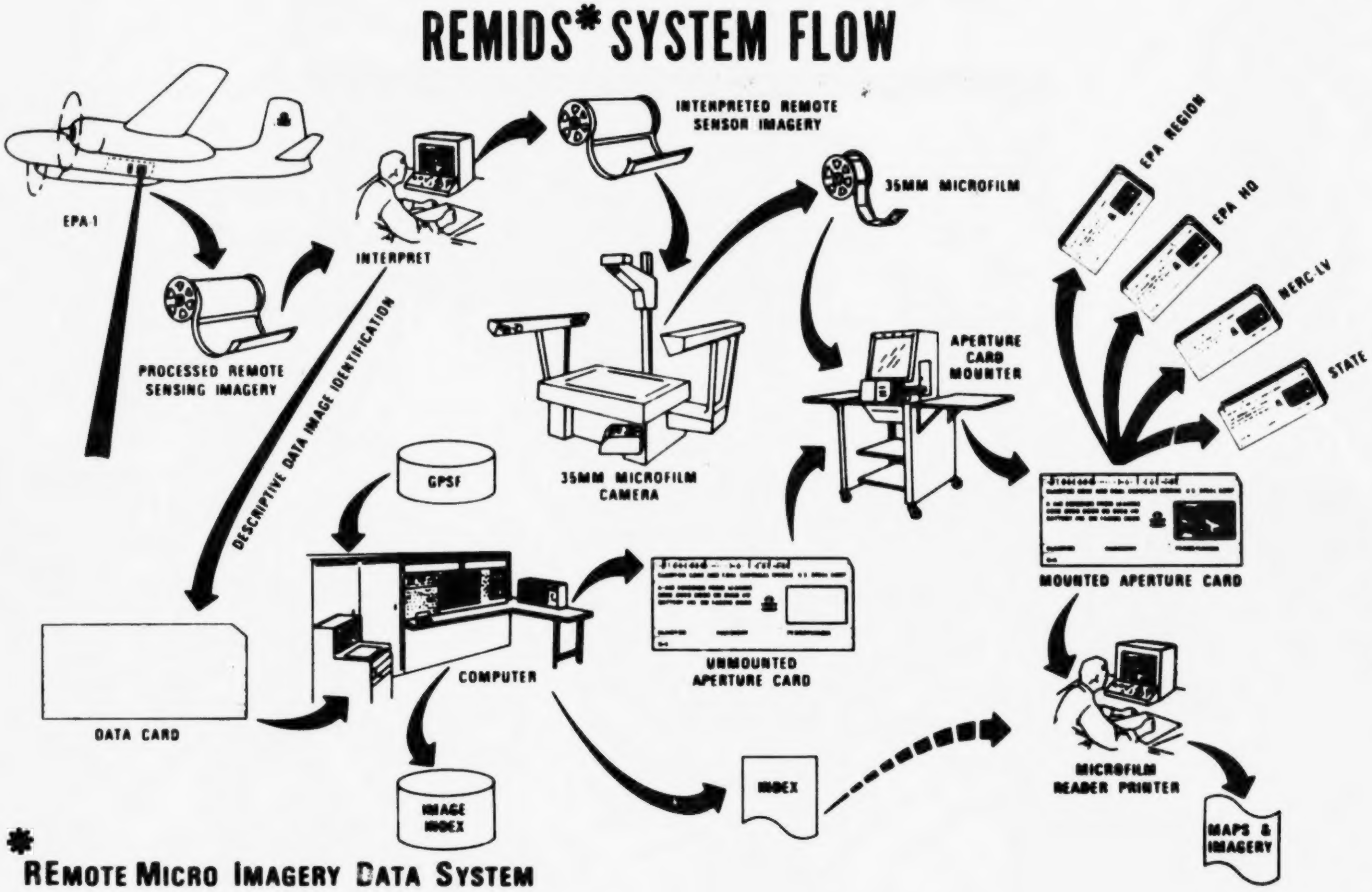
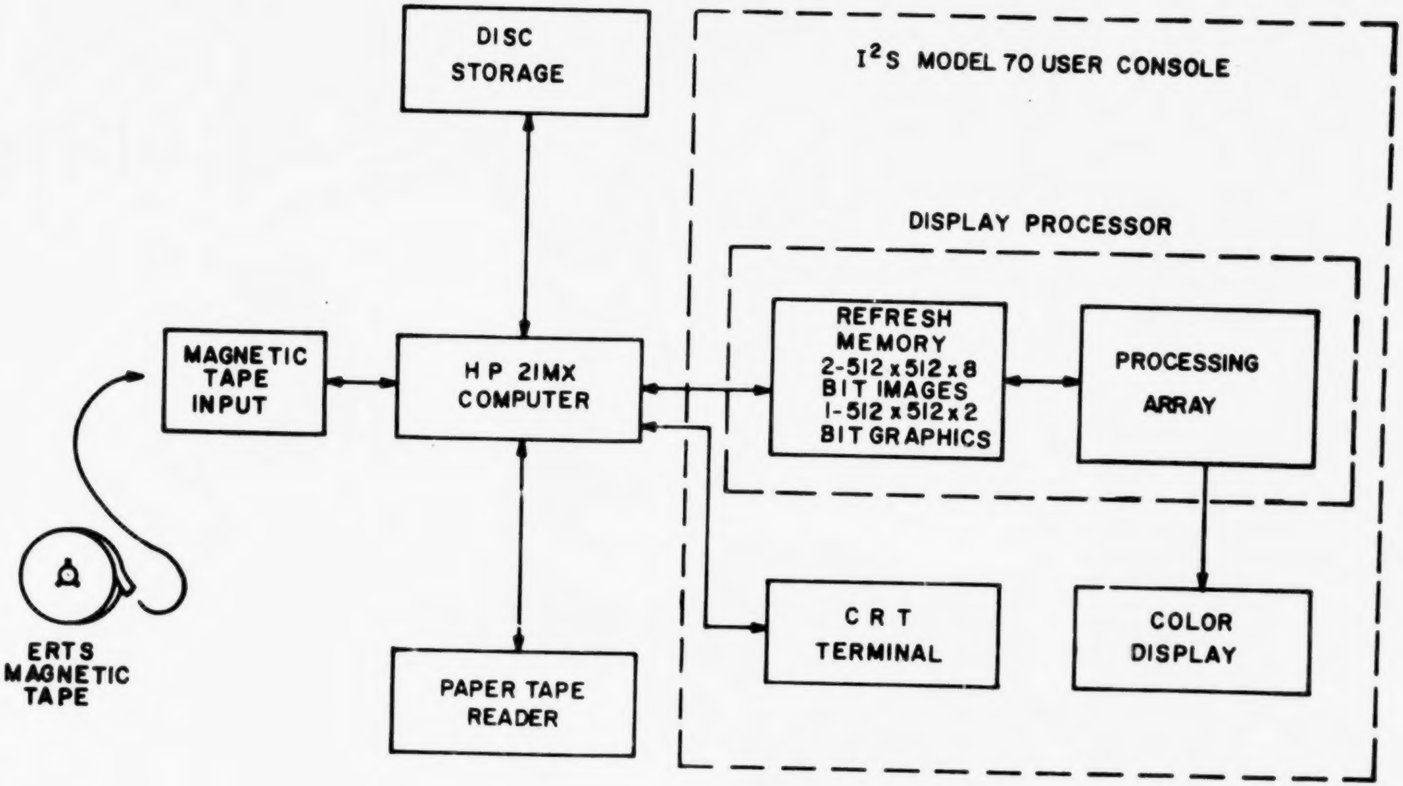

REMIDS* SYSTEM FLOW
EPA 1
PROCESSED REMOTE SENSING IMAGERY
INTERPRET
INTERPRETED REMOTE SENSOR IMAGERY
35MM MICROFILM CAMERA
35MM MICROFILM
APERTURE CARD MOUNTER
EPA REGION
EPA HQ
NERC-LV
STATE
MOUNTED APERTURE CARD
MICROFILM READER PRINTER
MAPS & IMAGERY
DESCRIPTIVE DATA IMAGE IDENTIFICATION
GPSF
DATA CARD
COMPUTER
IMAGE INDEX
UNMOUNTED APERTURE CARD
INDEX
* REmote Micro Imagery Data System

Figure 5


DISC
STORAGE
I²S MODEL 70 USER CONSOLE
DISPLAY PROCESSOR
MAGNETIC TAPE INPUT
HP 21MX COMPUTER
REFRESH MEMORY
2-512 x 512 x 8 BIT IMAGES
1-512 x 512 x 2 BIT GRAPHICS
PROCESSING ARRAY
ERTS MAGNETIC TAPE
CRT TERMINAL
COLOR DISPLAY
PAPER TAPE READER

MODEL 500
DIGITAL IMAGE PROCESSING SYSTEM

Figure 6

DEVELOPMENTS IN REMOTE SENSING PROJECTS

By Sidney L. Whitley

NASA's Johnson Space Center established the Earth Resources Laboratory (ERL) at the National Space Technology Laboratories (formerly Mississippi Test Facility) in late 1970 for conducting research investigations to develop applications of remote sensing. It was ERL's intention to use the large quantity of existing data acquired by aircraft and spacecraft as well as data to be collected in the future. ERL's mission statement is shown in Figure 1.*

ERL chose not to develop or refine sensors as other organizations in NASA were chartered for that purpose. The sensor technology was further advanced than user application technology and remains so today. It was not ERL's intention to develop data handling systems either; however, a data analysis capability was needed to develop applications of remotely sensed data. In 1971, ERL awarded a contract for the design and manufacture of a data analysis system. This system, known as the ERL-DAS, has been used to develop the applications shown in this paper. The ERL-DAS has also served as a test bed for the development of new low-cost data analysis systems which may be afforded by a larger number of potential remote sensor data users.

As an outgrowth of past contacts between individuals in EPA and NASA, a working agreement was recently established. This agreement was finalized in Memorandum of Understanding (MOU) D5-E771. The project resulting from this MOU is entitled, "Western Energy-Related Overhead Monitoring Project." Its application is directed toward monitoring the reclamation of strip mines in western United States. NASA has agreed to the following:

- To collect certain data with an airborne 11-band multispectral scanner (MSS) and a laser profiler
- To process selected NASA collected data
- To procure an 11-band MSS, a laser profiler, and a low-cost data analysis for EPA
- To train EPA personnel to use the equipment, associated software, and procedures.

Work began under this MOU in June 1975, and is progressing well. Figure 2 is a list of equipment and supplies NASA agreed to use in establishing a data acquisition and processing capability for EPA. It should be noted that the list includes the two sensors specified above, an image display system, a small computer, and several output recording devices. All work performed under this MOU is funded by energy pass-through funds.

In the course of ERL's research activities, several low-cost data analysis system (LCDAS) configurations have been defined. One of these configurations, low-cost data system configuration 4, closely matches the system specified in the EPA/NASA memorandum of understanding and is shown in Figure 3. EPA's LCDAS will be capable of reading data from EPA's airborne 11-channel MSS and laser profiler. The aircraft data will be recorded in Bi-Phase Level, Pulse Code Modulated (PCM) format. A PCM front-end will be added to the LCDAS shown in Figure 3 to allow the computer to read this highly specialized data format.

The EPA/LCDAS will be very similar to ERL's in-house data analysis system. A large number of data processing and applications programs will be provided to EPA under this MOU. EPA can adopt future ERL produced applications programs with little or no manpower expenditure because of the compatibility of our data analysis systems.

The ERL has developed and documented a large number of applications of remotely sensed data in our own research and in cooperation with other user agencies. Descriptions of some of these applications follow.

The ERL has developed a technique called the Water Search Program for detecting water vs. not-water. The results may be color or grey-shade coded on either a color film output or on an electrostatic printer/plotter. Figure 4 is an example of the water search output including a breakout of water and land area in acres. One application of this technique is shown in Figure 5. The technique is useful in studying the loss or buildup of shoreline, provided the data are carefully selected on different dates and at appropriate tide levels.

---

* Figures presented at the ADP Workshop were color photographs. This publication is limited to only black and white reproduction of these figures.

Much of our early spectral pattern recognition classification work was done in agricultural regions because ground truth is easily obtained and because fields are usually homogeneous. The technique is equally effective in studying marsh areas, such as the region shown in Figure 6. This particular marsh is known to be a salt marsh mosquito breeding area. Through the use of knowledge about the types of terrain on which mosquitoes breed, the types of vegetation that can exist on such terrain, and a vegetation classification map produced by spectral pattern recognition, one can infer a map which indicates potential for mosquito breeding. Figure 7 shows a salt marsh mosquito breeding map where red represents positive conditions, green represents negative conditions, blue represents water, and white represents other types of material, including roads, houses, and so forth.

Jointly, ERL and the National Marine Fisheries Service, both of NSTL, conducted a study to determine if menhaden fish catches could be related to water color. Through these studies, it was determined that menhaden fish were caught principally in waters of a certain color. A model was developed, LANDSAT data were input to the model, and a map of high, medium, and low potential for menhaden was produced. Figure 8 is an example of such a map.

A few months ago, ERL and the U.S. Army Corps of Engineers entered a study to determine if a certain Corps of Engineers-produced atlas could be updated with remote sensor data. It was determined that certain maps could be produced from LANDSAT MSS data. Figure 9 is a simulated color infrared photo map of the test area produced from MSS data. The map is composed of 27 LANDSAT frames collected in three seasons. The data has been translated from LANDSAT scene coordinates to the Universal Transverse Mercator (UTM) projection. The map was produced to a scale of 1:250,000, and the original product is accurate to about 300 meters, root mean square. The area shown in Figure 9 was also processed through ERL's spectral pattern recognition programs, and a surface classification of 24 material classes was produced. These 24 classes were aggregated to seven classes (i.e., individual crops were changed to agriculture, tree species were changed to forest, etc.), and a color-coded map was produced at 1:250,000 scale referenced to the UTM projection. Figure 10 is a photograph of the classified map produced by this technique.

The map shown in Figure 11 was produced by the same procedure as described above, but a greater number of categories were delineated in the map. It should be observed that the classification map has been superimposed over a quad map, and that the fit, which is particularly evident in the lower right corner of the figure, is quite good.

Certain regulatory agencies are quite interested in the extent of salt water intrusion in marshes. ERL botanists have used both aircraft and space acquired MSS imagery to survey salt marshes, brackish marshes, and fresh marshes. Figure 12 is an example of this capability, and another example of how well the remote sensor map can be made to fit the more conventional quad map.

During the past 3 to 4 years, ERL has greatly simplified and quickened its computer programs for processing remotely sensed multispectral scanner data, and has adapted these programs to run on small, widely available computers. During the past 2 years, inexpensive and highly capable image display systems have been designed and are now available commercially. Many users have a need for color-coded outputs of very high precision. The production of high quality color products has remained an expensive item.

ERL has conducted research to develop inexpensive techniques for color recording. Although the work is still in progress, there are some prelimiary results. One of the output devices ERL originally considered for production of grey shade maps is an electrostatic printer/plotter. This printer/plotter produces all of the standard line printer characters plus 16 shades of grey. It has been determined that the grey shade plots can be converted to color maps as described below.

The computer can be instructed to divide the digital imagery data into Red, Green, and Blue (RGB) components (or into separate land use material classes), or separate grey shade maps can be printed out for each component. These grey shade component maps are converted to film negatives and registered (the plotter is geometrically repeatable). The film positives can be converted to a color map using a $19.95 graphics kit plus an inexpensive black and white contact printer. The graphics kit, called Kwik Proof, was developed for the lithographic industry, and is used in proofing materials before an expensive lithographic run is made. The breakthrough needed was to format the scaled, digital image onto paper, and subsequently onto film so that the graphics kit could be used. Figure 13 is a color-coded soils map of Washington County, Mississippi, which was produced by this technique. There are nine soil types shown as different color levels in this scene. Only a very small portion of graphics materials was required to produce this product. The time required was about 2 hours, most of which was drying time. Figure 14 is a color-coded land-use map containing approximately 15 colors. This map shows that a large number of colors can be produced from only red, blue, and yellow (in this case) components. It should be observed that the computer superimposed coordinate lines registered quite well.

Another similar color output technique is under investigation by ERL. The required equipments are only slightly more expensive, but the product quality will be excellent and the processing time is short.

All of the techniques and applications will be available to EPA through the EPA/NASA agreement.

EARTH RESOURCES LABORATORY AT NSTL

MISSION

- CONDUCT RESEARCH INVESTIGATIONS IN MISSISSIPPI-LOUISIANA -GULF AREAS IN THE APPLICATION OF REMOTE SENSING.
- STRESS INTERESTS AND NEEDS OF AGENCIES IN THE AREA.
- UTILIZE EXISTING AIRCRAFT AND SATELLITE PROGRAMS AS A SOURCE OF DATA.
- COLLECT AND ANALYZE SURFACE DATA FOR CORRELATION WITH FLIGHT DATA.
- CONDUCT STUDIES OF USER REQUIREMENTS OF POTENTIAL APPLICATIONS IN ORDER TO GUIDE RESEARCH EFFORTS.

Figure 1

COST BREAKDOWN OF DATA ACQUISITION & PROCESSING HARDWARE

Image Display System		40K
FR2000 Tape Deck (Direct Read, All Speeds)		30K
FM Playback for Analog Recorded Data		30K
PCM Front End for Reading RS-18 Data		50K
Computer System		180K
V-74 Comp. w/32K MOS Memory	$38,400	
32K MOS Memory (Additional)	20,000	
Disc 46.7 M words	30,300	
Line Pr., 14" wide	10,200	
Statos, 33 Pr. Plotter, 22" wide	12,500	
2-120 IPS Tape Drives, 9-trk	16,000	
Tape Controller	6,000	
Card Reader	4,000	
Paper Tape Reader	2,300	
Floating Pt. Processor	5,000	
Expansion Chassis	1,000	
I/O Expander	600	
3-Buffer Interlace Controllers	1,500	
2-Priority Interrupt Modules	1,000	
1-Block Transfer Controller	1,500	
Cal Comp Plotter	30,000	
Film Recorder, B & W Strip, 5"		20K
Film Recorder, Color, Strip, Stand-Alone, 9.5"		120K
Auxiliary Air Conditioner, 15 tons for Computer Room		30K
Supplies (Tapes, Film, Paper, etc.)		25K
Support (Photo Lab, Printing, etc.)		20K
		$550K

11-Channel Airborne Multispectral Scanner	$200K
Airborne Terrain Profiler	50K
PCM Encoding System	70K
Set Ground Support Equipment, Cal, Checkout	75K
	$395K

Figure 2

LOW-COST DATA SYSTEM
CONFIGURATION 4

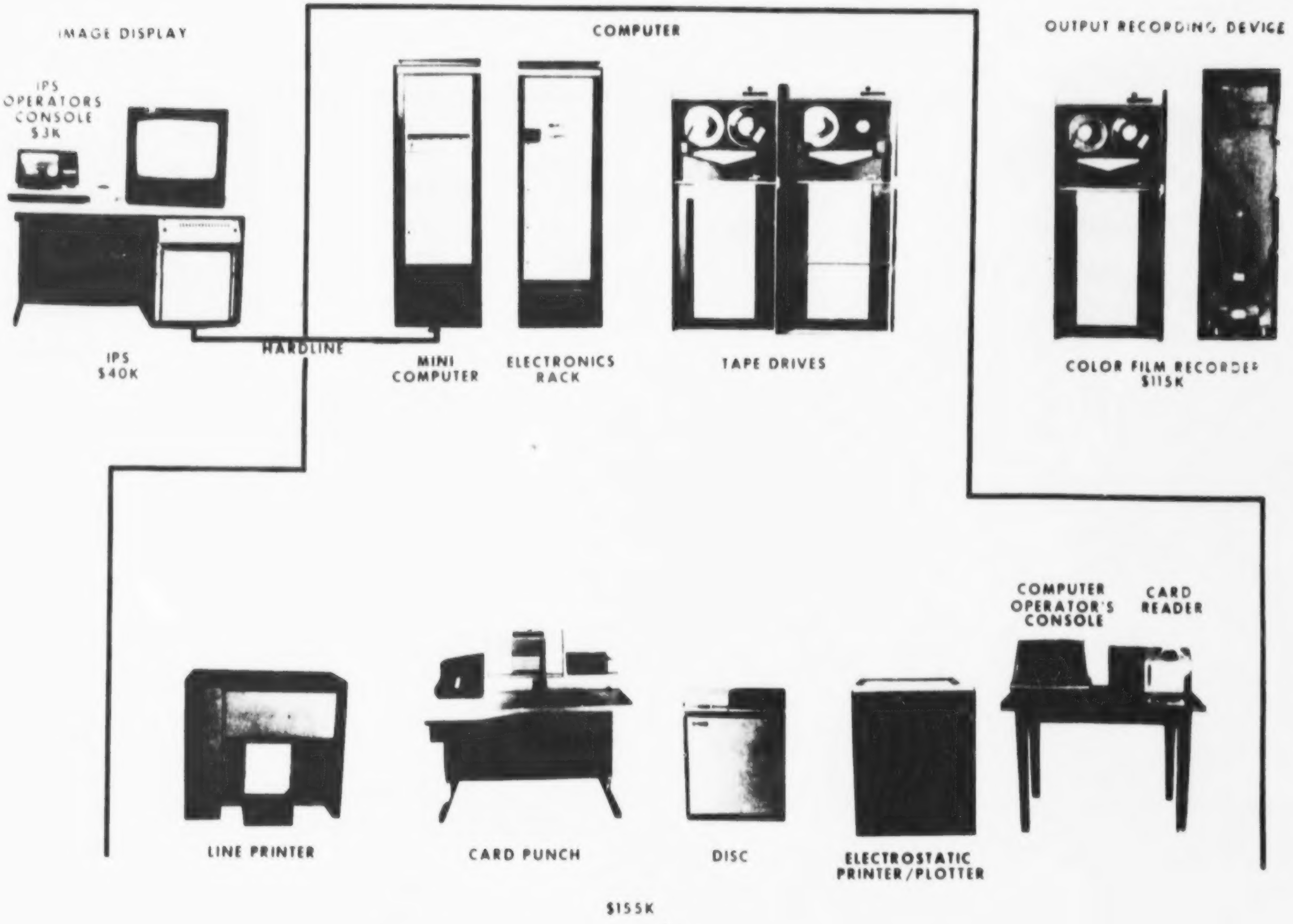
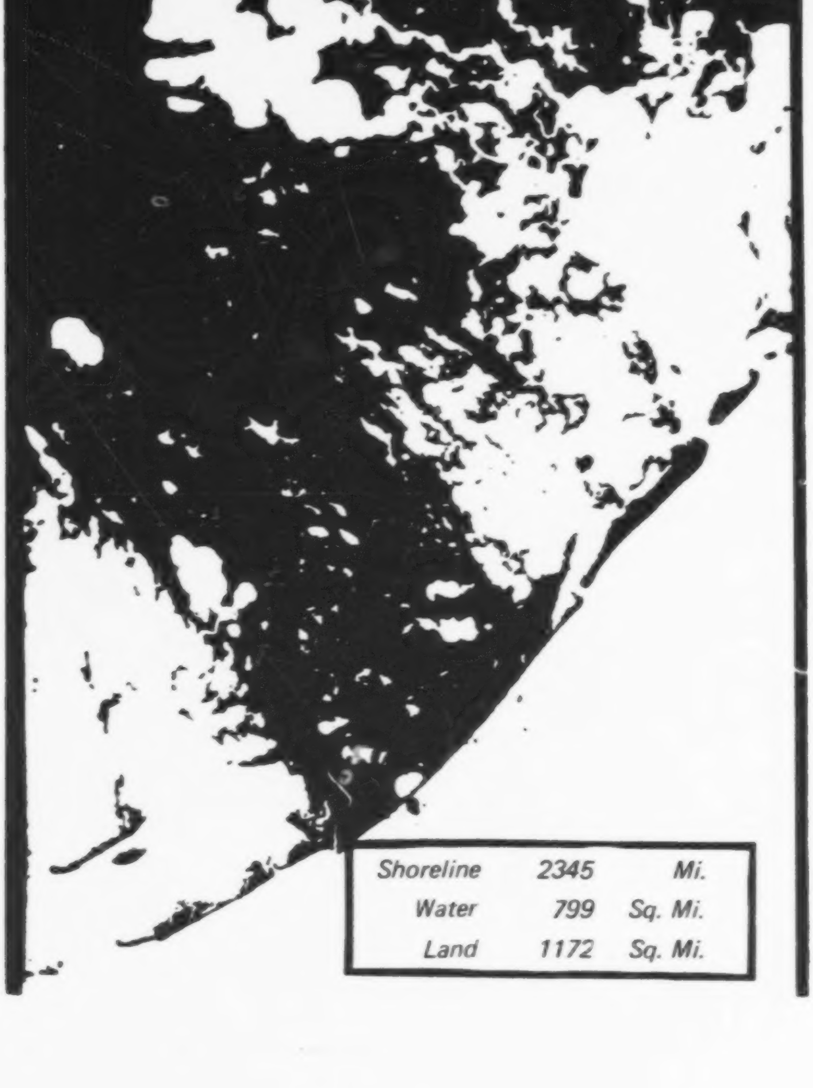


IMAGE DISPLAY
IPS OPERATORS CONSOLE $3K
IPS $40K
HARDLINE
COMPUTER
MINI COMPUTER
ELECTRONICS RACK
TAPE DRIVES
OUTPUT RECORDING DEVICE
COLOR FILM RECORDER $115K
LINE PRINTER
CARD PUNCH
DISC
ELECTROSTATIC PRINTER/PLOTTER
COMPUTER OPERATOR'S CONSOLE
CARD READER
$155K

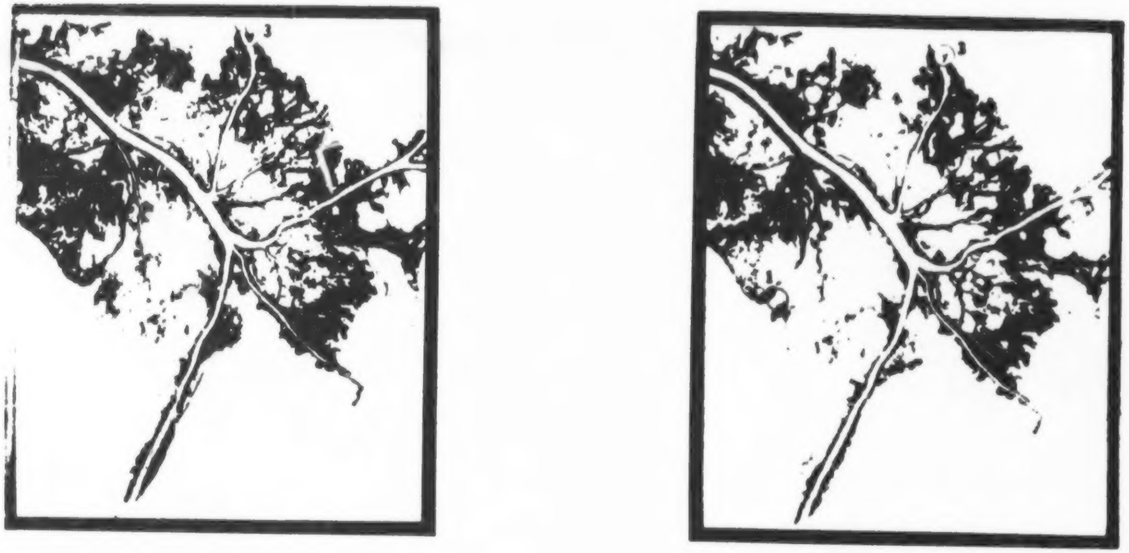
Figure 3



Shoreline	2345	Mi.
Water	799	Sq. Mi.
Land	1172	Sq. Mi.

Figure 4

LAND WATER INTERFACE ANALYSIS

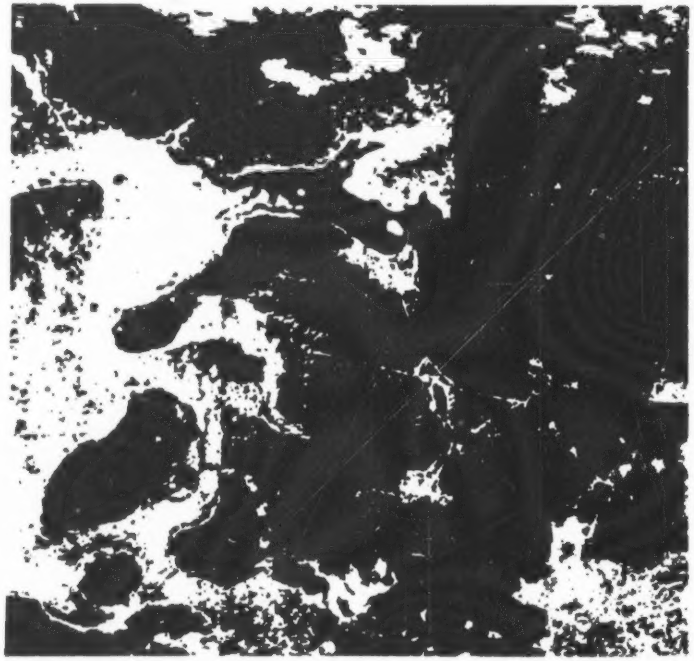


LANDSAT MSS DATA, PROCESSED BY WATER SEARCH AND SHORELINE ANALYSIS PROGRAMS INDICATES A 17 KM^2 LAND AREA DECREASE AND 58 KM LOSS OF SHORELINE AT MOUTH OF MISSISSIPPI RIVER, BETWEEN JANUARY & DECEMBER OF 1973

Figure 5

◇ MARSH ECOLOGICAL STUDIES ◇

COMPUTER DERIVED ECOTYPES



ERL MTF

CLASSIFICATION

- ■ *SPARTINA PATENS* & *JUNCUS ROEMERIANUS*
- ■ TREES & SHRUBS
- ■ OPEN WATER
- SAWGRASS (*CLADIUM JAMAICENSE*)
- ■ WHITE WATER-LILY (*NYMPHAEA ODORATA*)
- CATTAILS (*TYPHA* SP.)
- *ELEOCHARIS QUADRANGULATA*
- UNIDENTIFIED *GRAMINEAE*

Figure 6

MARSH ECOLOGICAL STUDIES

FRITCHIE MARSH, WHITE KITCHEN, LA

SALT MARSH MOSQUITO (AEDES SOLLICITANS, WLK)
BREEDING MAP (INFERRED FROM VEGETATIONAL
CLASSIFICATION MAP)

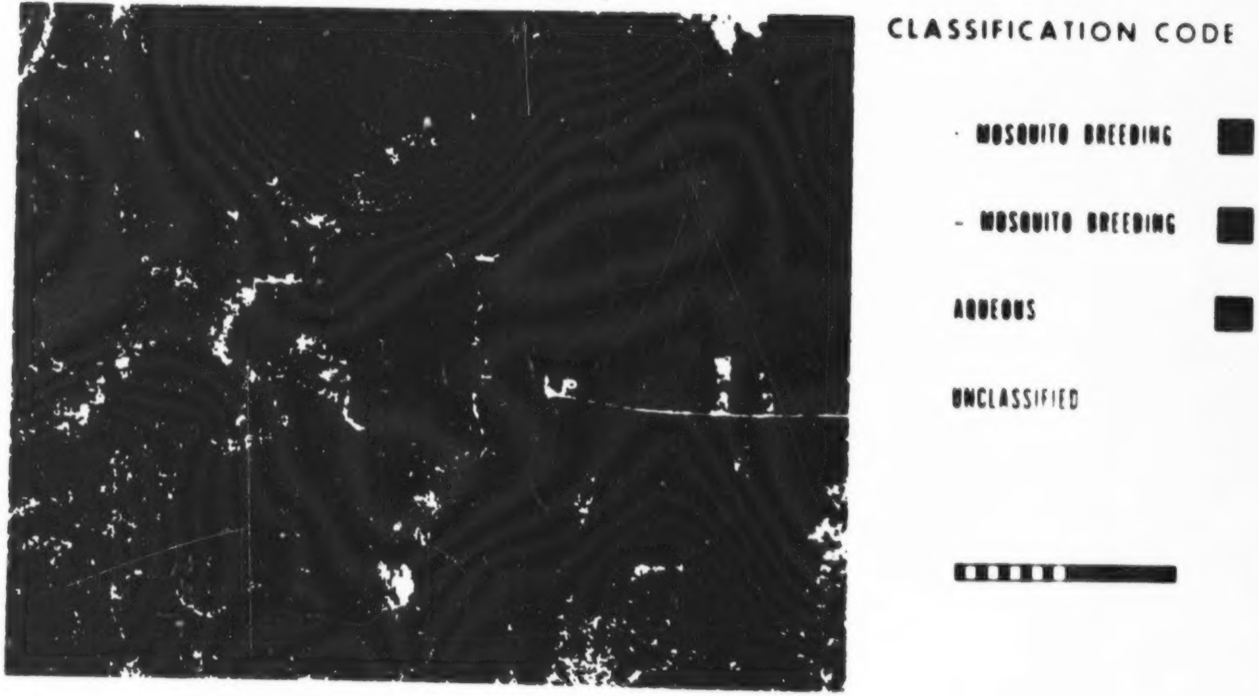

CLASSIFICATION CODE
· MOSQUITO BREEDING ■
- MOSQUITO BREEDING ■
AQUEOUS ■
UNCLASSIFIED

Figure 7

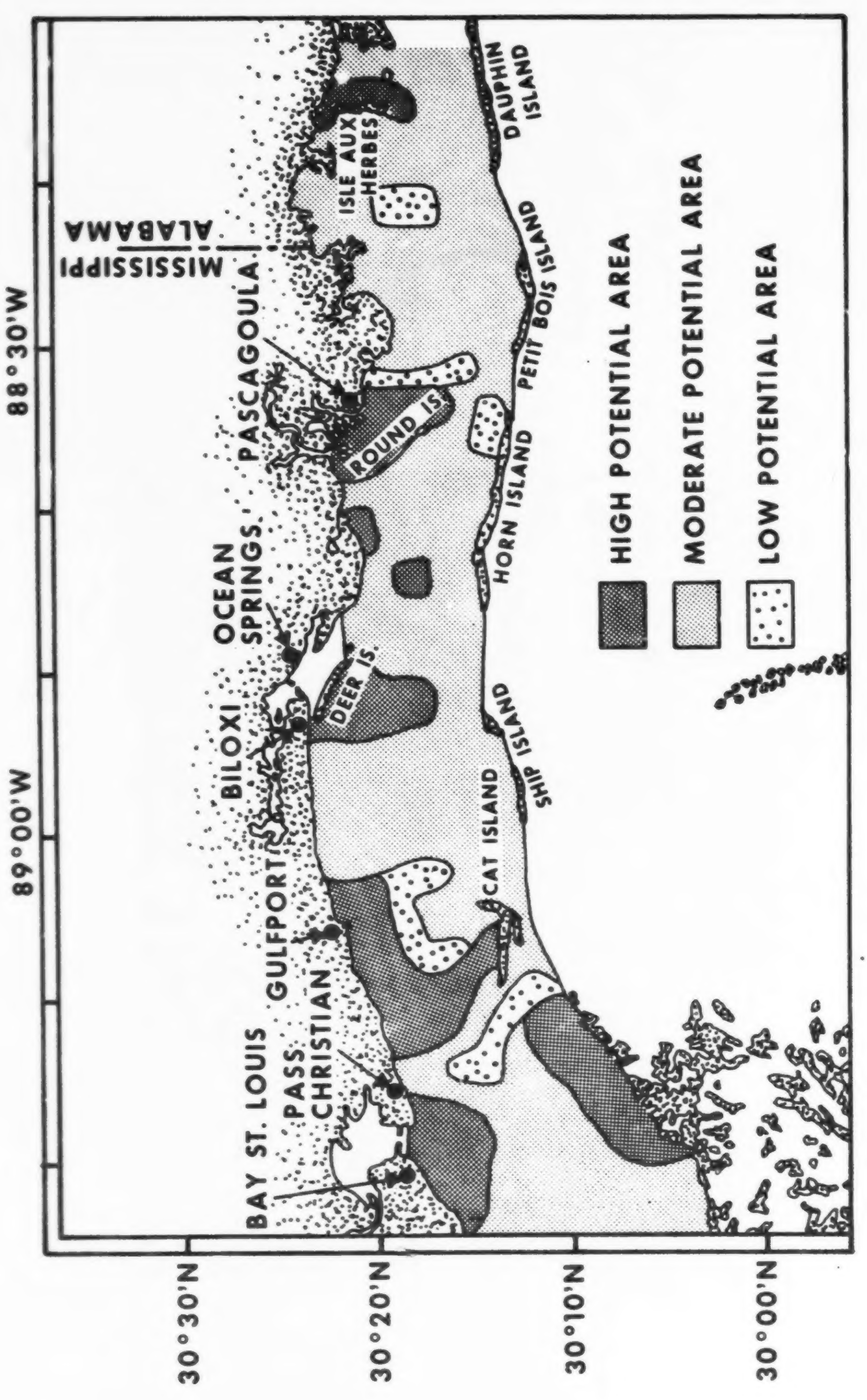

89°00'W
88°30'W
30°30'N
30°20'N
30°10'N
30°00'N
BAY ST. LOUIS
PASS CHRISTIAN
GULFPORT
BILOXI
OCEAN SPRINGS
PASCAGOULA
MISSISSIPPI
ALABAMA
DEER IS.
ROUND IS.
ISLE AUX HERBES
CAT ISLAND
SHIP ISLAND
HORN ISLAND
PETIT BOIS ISLAND
DAUPHIN ISLAND
HIGH POTENTIAL AREA
MODERATE POTENTIAL AREA
LOW POTENTIAL AREA

Figure 8

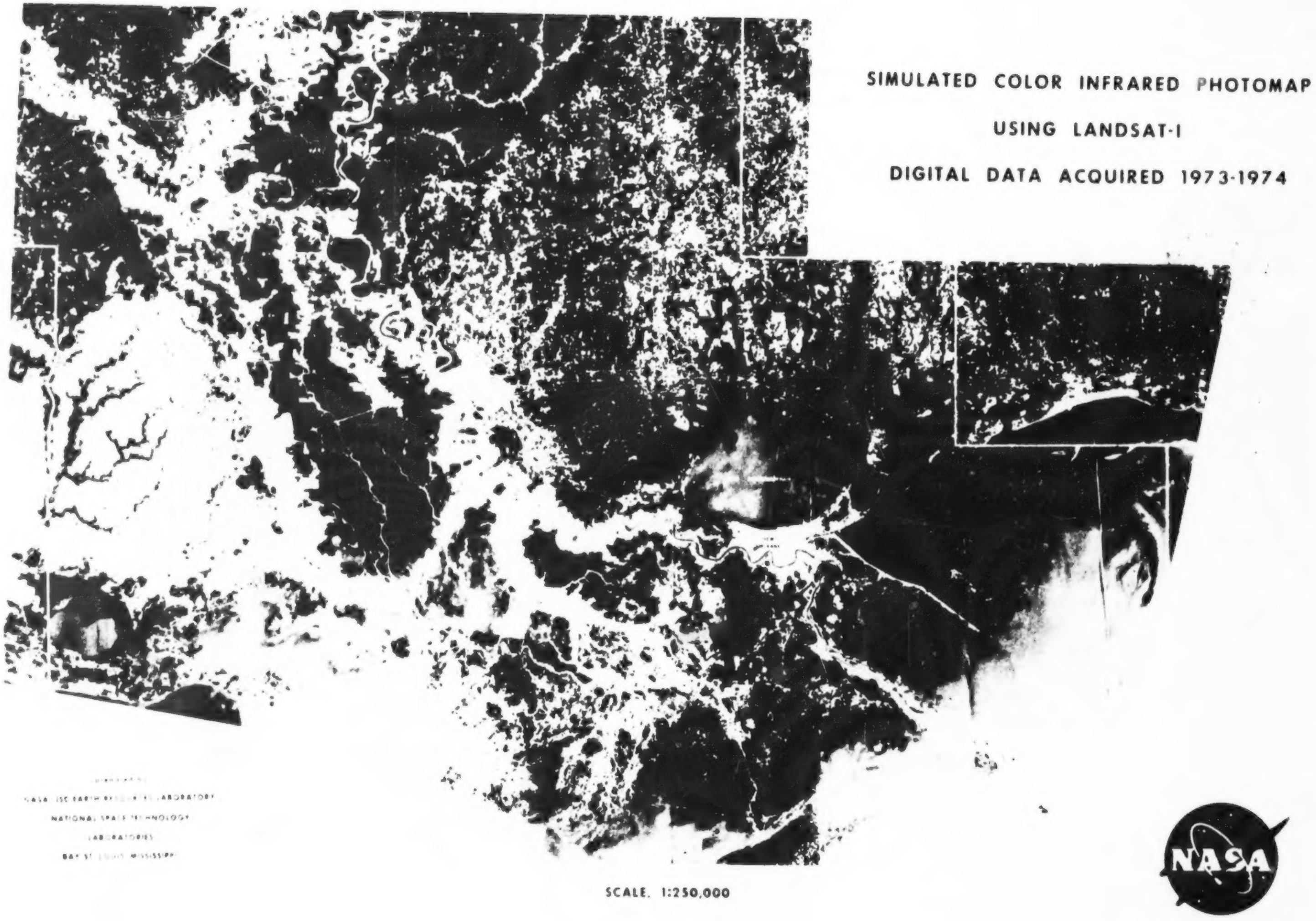

SIMULATED COLOR INFRARED PHOTOMAP
USING LANDSAT-1
DIGITAL DATA ACQUIRED 1973-1974
NASA JSC EARTH RESOURCES LABORATORY
NATIONAL SPACE TECHNOLOGY
LABORATORIES
BAY ST. LOUIS MISSISSIPPI
SCALE, 1:250,000

Figure 9

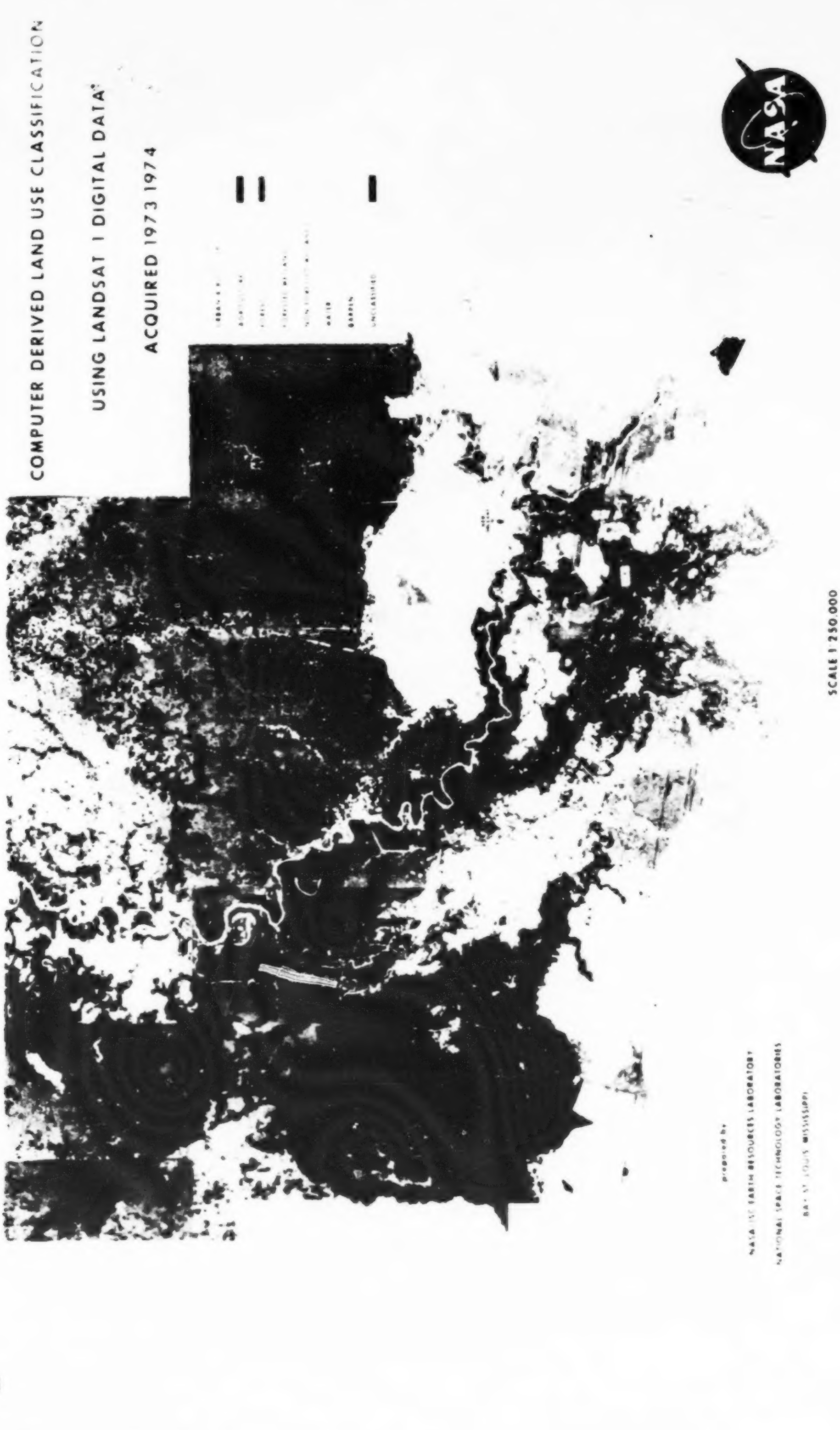

COMPUTER DERIVED LAND USE CLASSIFICATION
USING LANDSAT 1 DIGITAL DATA
ACQUIRED 1973 1974
WATER
BARREN
UNCLASSIFIED
SCALE 1:250,000
prepared by
NASA JSC EARTH RESOURCES LABORATORY
NATIONAL SPACE TECHNOLOGY LABORATORIES
BAY ST. LOUIS, MISSISSIPPI

Figure 10

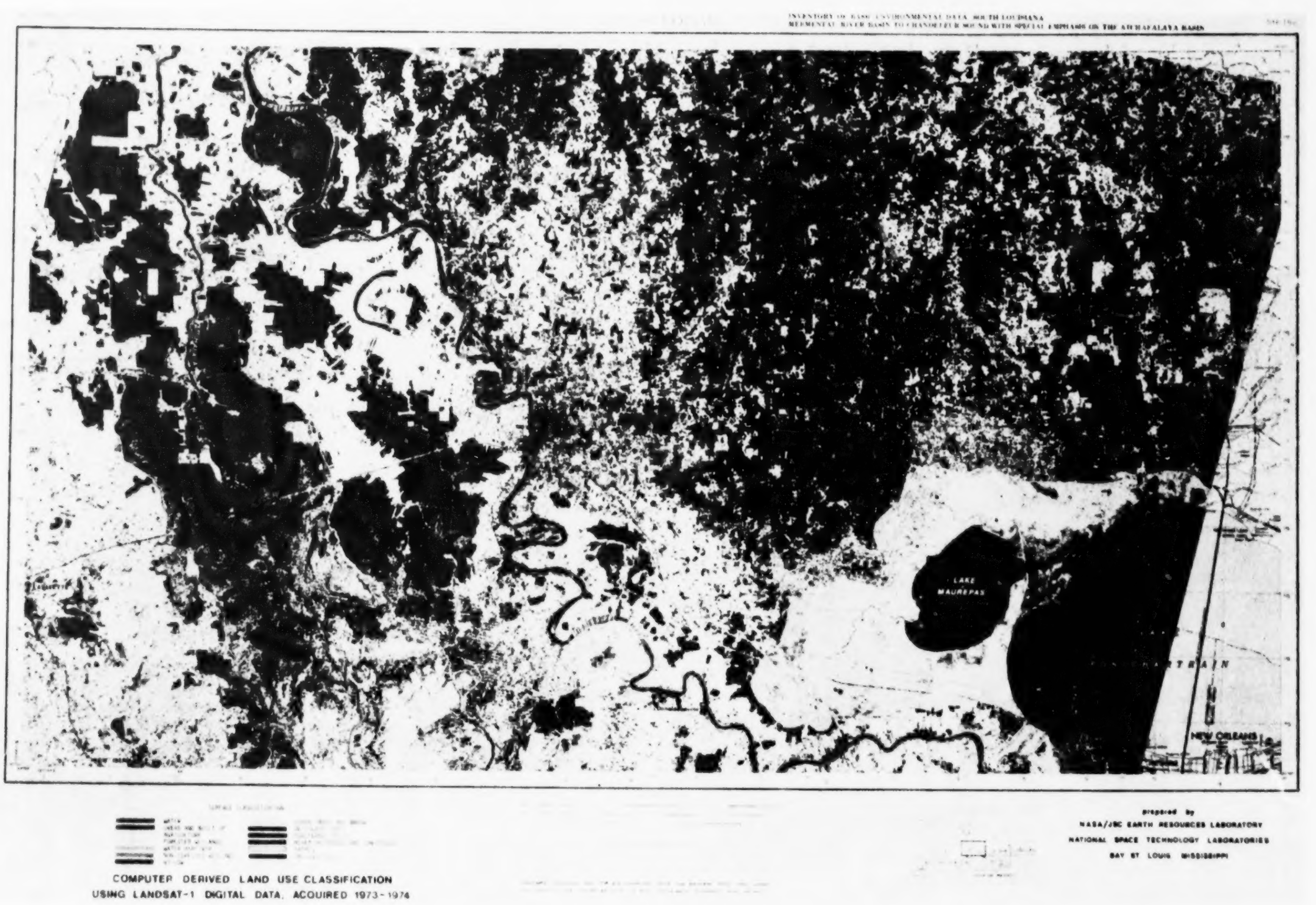

INVENTORY OF BASIC ENVIRONMENTAL DATA, SOUTH LOUISIANA
MERMENTAU RIVER BASIN TO CHANDELEUR SOUND WITH SPECIAL EMPHASIS ON THE ATCHAFALAYA BASIN
LAKE MAUREPAS
NEW ORLEANS
COMPUTER DERIVED LAND USE CLASSIFICATION
USING LANDSAT-1 DIGITAL DATA, ACQUIRED 1973-1974
prepared by
NASA/JSC EARTH RESOURCES LABORATORY
NATIONAL SPACE TECHNOLOGY LABORATORIES
BAY ST LOUIS MISSISSIPPI

Figure 11

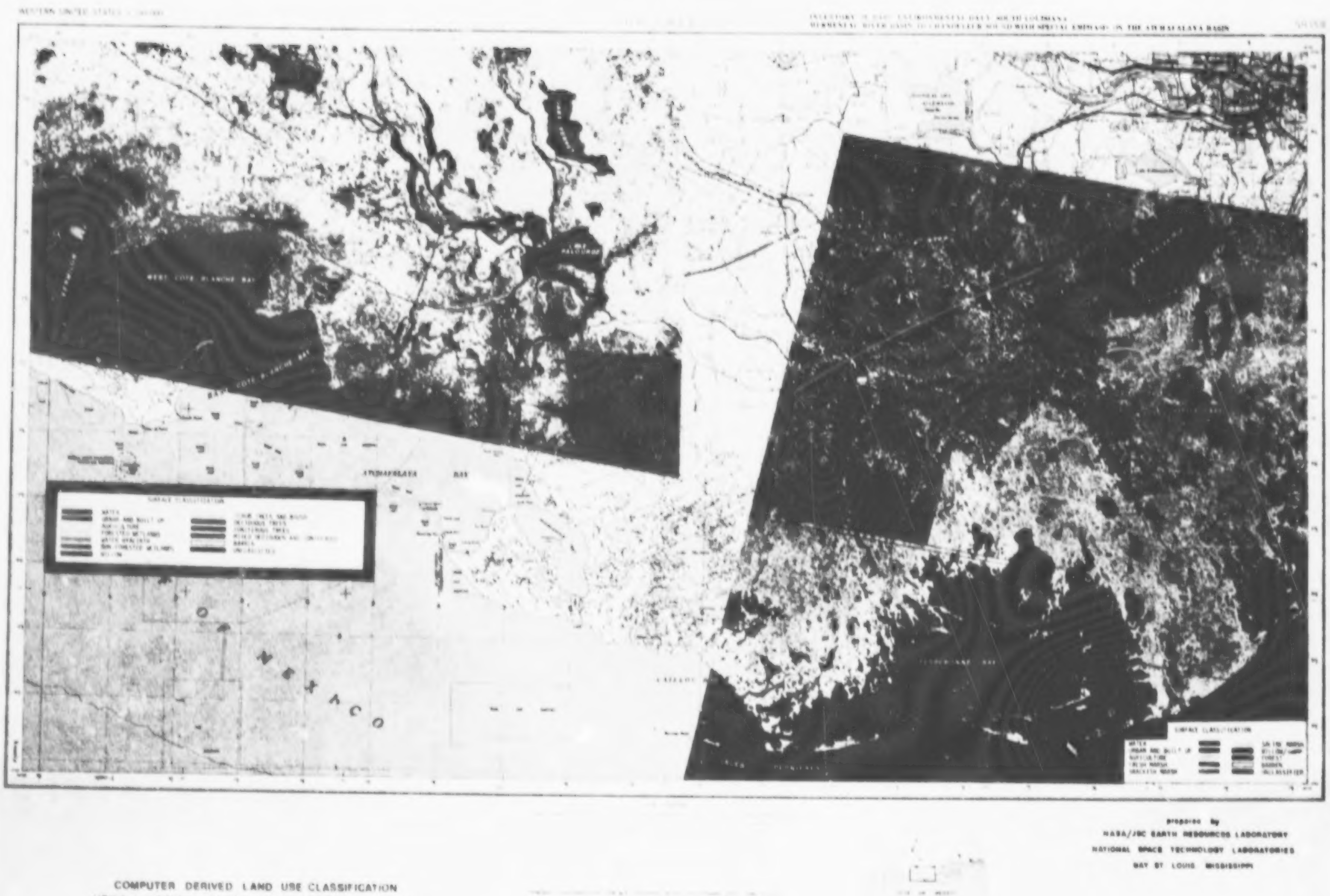
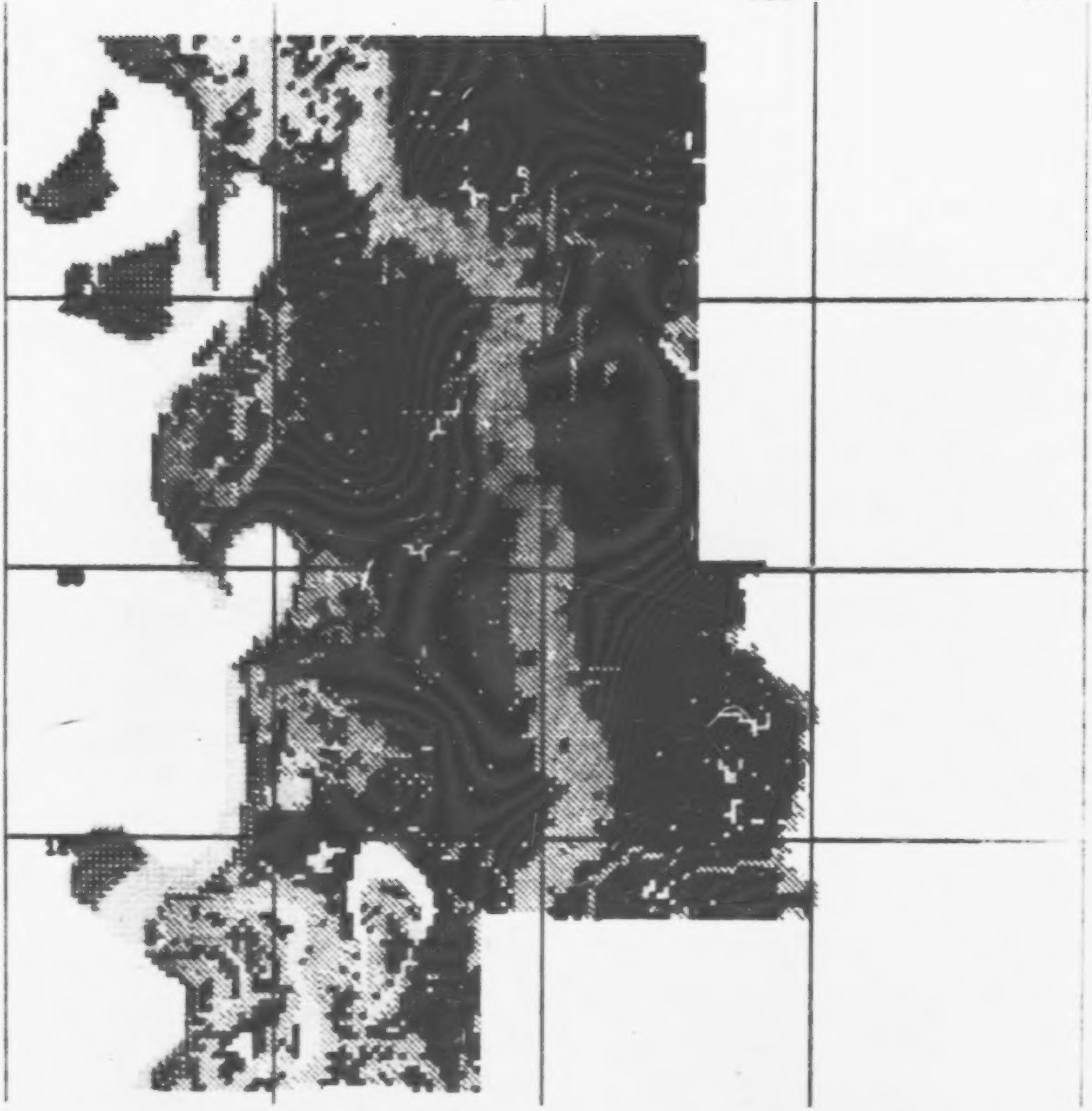

INVENTORY OF BASIC ENVIRONMENTAL DATA, SOUTH LOUISIANA
MERMENTAU RIVER BASIN TO CHANDELEUR SOUND WITH SPECIAL EMPHASIS ON THE ATCHAFALAYA BASIN
WEST COTE BLANCHE BAY
EAST COTE BLANCHE BAY
ATCHAFALAYA BAY
SURFACE CLASSIFICATION
WATER
URBAN AND BUILT UP
AGRICULTURE
FORESTED WETLANDS
WATER HYACINTH
NON FORESTED WETLANDS
DECIDUOUS TREES
CONIFEROUS TREES
MIXED DECIDUOUS AND CONIFEROUS
BARREN
UNCLASSIFIED
O
M
E
X
C
O
SURFACE CLASSIFICATION
WATER
URBAN AND BUILT UP
AGRICULTURE
FRESH MARSH
BRACKISH MARSH
FOREST
BARREN
UNCLASSIFIED
prepared by
NASA/JSC EARTH RESOURCES LABORATORY
NATIONAL SPACE TECHNOLOGY LABORATORIES
BAY ST LOUIS MISSISSIPPI
COMPUTER DERIVED LAND USE CLASSIFICATION
USING LANDSAT-1 DIGITAL DATA, ACQUIRED 1973-1974
13

Figure 12



9 LEVEL DIGITIZED SOILS MAP PRODUCED FROM 3 GRAYSCALE PLOTS FROM ELECTROSTATIC PRINTER PLOTTER

PROCESS REQUIRES $19.95 GRAPHICS KIT
PLUS SMALL B & W CONTACT PRINTER ($600 to $2,000)

Figure 13

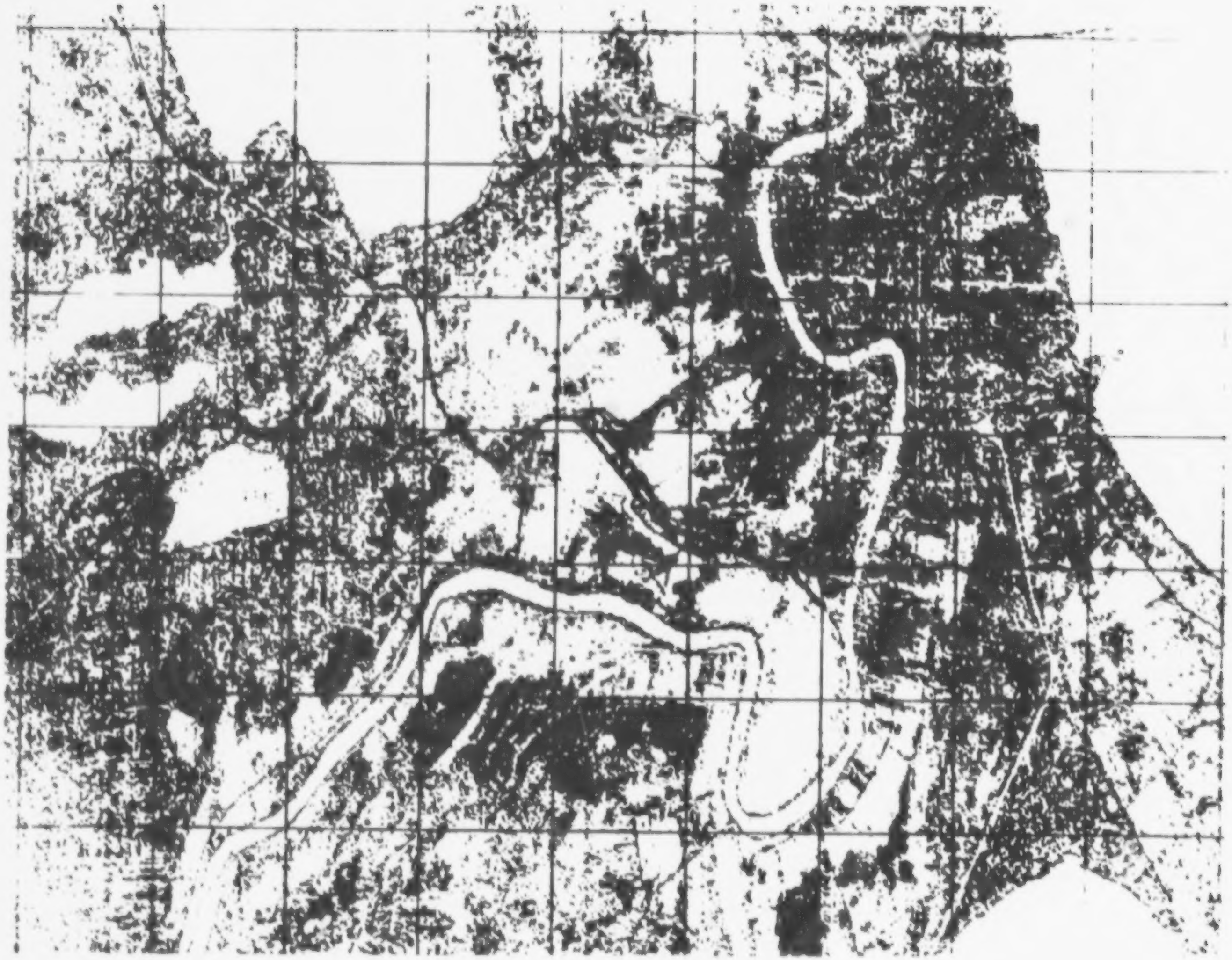


Figure 14

SUMMARY OF DISCUSSION PERIOD - PANEL V

This discussion period, on the Developments in Remote Sensing Projects, included the following remarks.

Definition of Remote Monitoring

Remote monitoring is defined by EPA as including not only remote sensing which uses instruments such as lasers and multispectral scanners, but also automated in situ contact monitoring platforms in which the information is telemetered back to some central location.

CHAMP System

The difficulties experienced in the CHAMP system were discussed.

The panel felt that the original specifications were not too loose and the data system had no major weaknesses. Documentation of system programs was completed slowly, however. It was agreed that the main weakness can be traced to the instrumentation. Although most instruments met initial specifications, they were not thoroughly field tested before being employed. RAPS encountered similar problems. If the system were to be implemented again from the start, a stronger emphasis should be placed on field maintenance. Problems with instrumentation were handled as brushfires. More flexibility should have been introduced into the RAPS system initially to handle such instrument problems. The question of how to handle data that can be accurately adjusted for known instrument error must be addressed. It should be decided whether to flag the data as invalid, or to correct the data and document the changes with appropriate comments. Emphasis should be placed on system requirements instead of instrument manufacturers' specifications.

A considerable portion of the CHAMP error checking is done manually. It was asked whether more could have been done by machine. Eventually all will be done by machine. About 85 percent of the data is now completely machine-validatable. Only quality spot checking should be required.

NASA-ERL System

The applications which are planned as part of the EPA low cost data analysis system were discussed. The software developed by NASA-ERL was designed predominantly for monitoring the reclamation of strip mining activities in western United States, but the software applications are available to EPA. As the system is transmitted, training sessions will be coordinated. The system being developed under a work agreement with EPA can now be used in a hands-on environment at the NASA-ERL facility.

Monitoring Techniques

The technique used for discrimination between smog, smoke, and fog was explained. Interpretation can be made by coupling the presence of smog, smoke, or fog with concomitant happenings in the surroundings; i.e., stacks and meteorologic conditions. Color analysis can also be used.

It was asked whether ultraviolet (UV) fluorescence is specific for petroleum. Natural organics can be picked up if they fluoresce. Lasers detect not only oil but also its thickness. They also give some information as to the type of oil.

EPA will cope with offshore monitoring platforms and with monitoring the Continental Shelf as follows. The Office of Monitoring and Technical Support must first determine what tools are needed for the offshore oil rigs. Within 10 years, there will be about 1,000 large platforms off the coast of the United States. Industry will, of course, try to cut costs. It is hoped that through proper prioritization of resources within ORD, sensors will be developed that are not only overhead monitoring types, but data buoys which can be put in select locations to give profiles of exactly what is happening. Interagency agreements will be needed to accomplish this task.

Storing Remote Sensing Data

The best system for storing volumes of remote sensing data depends on what the data uses will be. Conventional data bases are inadequate to handle the large data bases that result from remote sensing. One system, the REMID system, uses the computer as an index into the data base.

Commercial Software Use

The contour package used in Las Vegas is a commercial software package that is in use. It was developed by a small company and is very similar to the conventional plotting routines.

Pollution Monitoring

Remote sensing data does not include information for pollutants covered by environmental standards, but airborne remote sensing is not a wasted luxury. Information is required by the Agency besides the pollutant concentration at specific points. There are already examples of court cases in which remote sensing data were used as evidence. Eventually, remote sensors will be developed for specific pollutants. Remote sensing will give information on where pollutants are concentrated and where further monitoring should be performed by in situ monitoring. Remote sensing is also being used by Region V to determine pollution sources and for nonpoint source pollution monitoring.

EPA Precedence in Remote Sensing

To get management acceptance and Agency utilization of remote sensing, the Agency must reprioritize. Instead of waiting for other agencies to lead the way, EPA must take the lead in certain areas.

AGENCY NEEDS AND FEDERAL POLICY

By Melvin L. Myers

This presentation will consider:

- Where EPA currently stands and where it is heading
- Trends within EPA
- Problems which EPA encounters
- Functions which EPA needs to perform and resulting data base requirements
- Federal constraints on the usage of ADP systems.

As the Administrator has remarked, EPA should be examining accomplishments over the last 5 years and structuring programs for the next 5 years.

Future trends in where EPA is going include:

- Defining EPA's goals and objectives for the next 5 years in addition to compliance with statutory deadlines
- Concentrating on preventing environmental deterioration as well as abating pollution
- Strengthening our Federal, State, and local partnership in environmental programs.

A possible framework for structuring these trends within the Agency is illustrated in Figure 1; a strategic approach to defining goals and a waste management approach to deterioration prevention. The figure highlights implementation as the output of our environmental programs.

The EPA has been encountering several problems, including:

- Difficulty in court cases
- Efficacy in meeting our environmental standards
- Shifting efforts to maintenance of clean air
- Institutional demonstration of waste management practices
- The question of our role in radiation and that of the Nuclear Regulatory Commission
- The implementation of areawide planning
- The degree of regulating pesticides
- The legal problems in proving adequate quality assurance.

The Agency is addressing conditions of its own management environment, including:

- Economic, energy, and environmental impacts of regulations
- Emerging toxic substances legislation
- Increased Regional/headquarters interaction in enforcement
- Firm technical backup developed in support of regulatory action
- Novel approaches such as our fuel economy program
- Requirements of the Freedom-of-Information Act.

The Agency recognizes the need for ADP policy and has established a steering committee to develop an Agency 5-year ADP plan. The committee is comprised of the five Assistant Administrators.

Agency functions will be used as a basis for establishing the need for current or new data bases. Figure 2 shows how this may be done through an input-output matrix and within the pattern of the functions as illustrated in Figure 1.

A list of Agency functions for which data bases are and will be required include the following (see Figure 2):

- Administration (personnel, finance)
- Implementation (enforcement, citizen participation, monitoring)
- Strategic analysis (safe environment, clean environment)
- Resource recovery (reuse, reclamation, and recycling of energy and materials)
- Deterioration prevention (areawide planning)
- Consumption modification (fuel economy, waste paper source reduction)
- Production modification (alternative input, processes, and practices)
- Pollution abatement (ambient standards, source standards)
- Usage restrictions (pesticide classifications)
- Research (toxicology, pollutant characterization, research monitoring)
- Development (prototype technology, model validation)
- Demonstration (control technology, alternative technology)
- Quality assurance (pollution and effects measurement, standards achievement, monitoring planning).

These functions require many special considerations in the development of data bases. We need good documentation from which to respond to Freedom-of-Information Act requests, and we need to question whether large systems or small systems should be used. We need a national communications network, a way to secure trade secrets when registering products, a method to allow for professional peer review of data, and a total Agency approach of maintaining valid information on the status of environmental quality and its relationship to meeting our standards.

The Office of Management and Budget (OMB) must oversee Federal use of ADP resources. The relationship between OMB, the General Services Administration (GSA), and the other Federal Agencies is usually misunderstood. Actually, the significance, both domestically and internationally, of the Federal influence on the ADP market warrants overseeing by OMB. Although OMB would prefer to limit its role to program budget decisionmaking, it must, in the case of ADP, also act as a guardian against proliferation and misuse of a means to an end.

The OMB function is not a regulatory one, but because of the implications of Federal ADP policy, it maintains a veto right over GSA and other agencies concerning the usage of computers. OMB relies upon GSA for policy direction and project impact analysis, as well as for computer-related alternatives, costs, and benefits. OMB may get involved when implementation is not consistent with its policy.

Over the last 3 to 4 years, OMB has consciously implemented its policy; its hypothesis is that the first avenue should be to use the private sector to fill an ADP need. In-house systems resulting in utilities should be avoided.

There are two primary justifications for remaining in-house for computer systems. One reason is if it is clearly to the Nation's advantage, and the other is cost.

If a computer is to be provided in-house, then three stages must be considered. First, is there surplus time on another Government computer? Second, can excess equipment be reused to fill the need? Third, should the equipment or service be rented or purchased?

OMB has urged GSA many times to change its policy regarding minicomputers. Current policy prefers larger systems, but we need smaller systems.

Implementation of GSA's Federal Schedule 66, Laboratory Instruments and Automation, is hopefully, lessening administrative resistance.

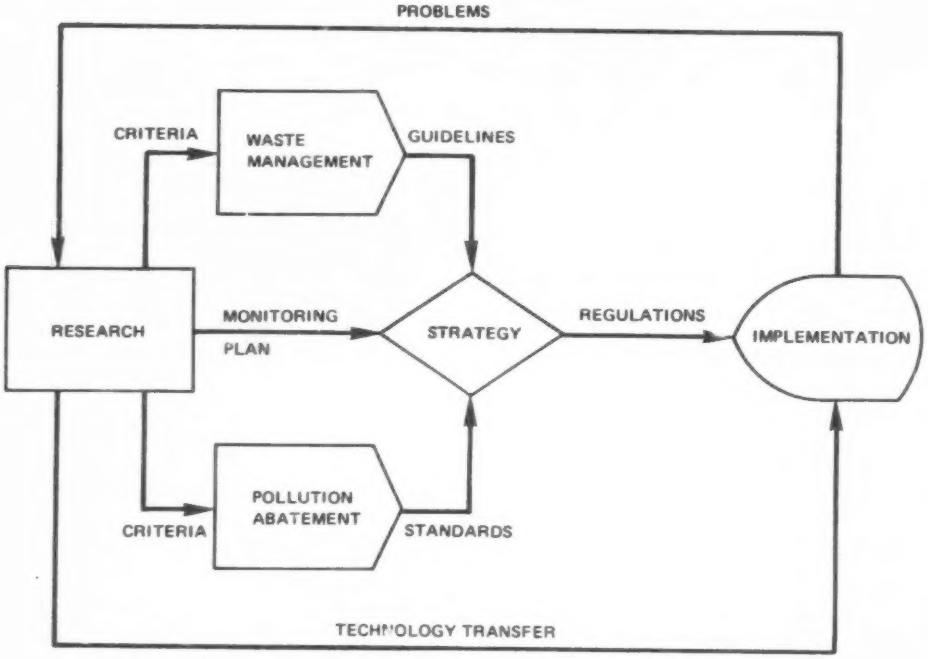

PROBLEMS
CRITERIA
WASTE MANAGEMENT
GUIDELINES
RESEARCH
MONITORING PLAN
STRATEGY
REGULATIONS
IMPLEMENTATION
POLLUTION ABATEMENT
CRITERIA
STANDARDS
TECHNOLOGY TRANSFER

Figure 1
A Functional Pattern of the Agency System

		FILE		INPUT		OUTPUT		
	FUNCTION	PERMIT COMPLIANCE SYSTEM	FUELS DATA BASE	PERMITS	LAB ANALYSIS DATA	TREND REPORTS	FACILITY REPORT	COMPLIANCE SCHEDULES
	ADMINISTRATION							
	IMPLEMENTATION			●	●		●	●
	STRATEGIC ANALYSIS							
WASTE MANAGEMENT	RESOURCE RECOVERY							
	DETERIORATION PREVENTION					●	●	
	CONSUMPTION MODIFICATION					●		
	PRODUCTION MODIFICATION					●		
POLLUTION ABATEMENT	POLLUTION ABATEMENT						●	●
	USAGE RESTRICTIONS					●		
RESEARCH	RESEARCH					●		
	DEVELOPMENT							
	DEMONSTRATION							
	QUALITY ASSURANCE				●	●		
FILE	PERMIT COMPLIANCE SYSTEM			●			●	●
	FUELS DATA BASE				●	●		

Figure 2
A Partial Input-Output Matrix Approach Relating Agency Functions to System Files

MINICOMPUTERS: CHANGING TECHNOLOGY AND IMPACT ON ORGANIZATION AND PLANNING

By Edward J. Nime

Because of recent developments in computer electronic technology, computing power will soon be essentially free. Since 1968, hardware costs of the minicomputer classes have been decreasing at a rate of approximately 30 percent each year, with an equivalent increase in processing capability.

The Environmental Protection Agency (EPA) will benefit from this trend, which will allow many of the functions currently performed on large central computers to be performed locally in Regional centers. There are potential organizational and management problems associated with the use of this technology; however, with effective management coordination and control, EPA can maximize present benefits and allow for easy expansion or integration in the future.

The laboratory automation workshop presentations represent developments made possible by recent breakthroughs in electronics technology. It is generally known that the first generation of computers contained vacuum tubes, the second generation, transistors, and the third generation, integrated circuits. Large scale integration (LSI) has resulted in a computer-on-a-chip; *i.e.*, a general purpose, programable integrated circuit equivalent to a central processor unit of a conventional computer, on a chip of silicone a few millimeters on a side and containing the equivalent of several thousand transistors.

These technology developments are resulting in micro- and minicomputers which approach the processing capabilities of medium-sized systems like the IBM 370/135, but at an extremely small fraction of the cost. Since 1968, minicomputer costs have been decreasing approximately 30 percent per year with an equivalent increase in processing capability. A processor which cost $25,000 in 1966 can be purchased for less than $2,000 today. Nearly 80 percent of all minicomputers have a processor in the price range of $2,000 to $10,000.

A 1974 Auerbach study states that the number of computer installations in existence today represents less than 5 percent of the total computing power projected for 1984. Last year, minicomputers represented $1.2 billion or 13 percent of the total computer market; and by 1984, the total hardware market will have increased to $20 billion with minicomputers accounting for $6 billion or 30 percent of the total.

A question often asked is "Will minicomputers replace large computers?" The answer is "No." There will always be a need in EPA for large computers like an IBM 370/158 or a Univac 1110 which can process great masses of water and air quality data for trend analyses and modeling. Likewise, there will always be a need for dedicated and general purpose minicomputers which bring easy-to-use, dependable, responsive, and cheap computing power to many. All EPA computer application areas can benefit from minicomputers, especially when linked to large computers.

A relatively new term which describes the shared processing of minicomputers and large computers is "distributed processing." In the context of this discussion, distributed processing is the interconnection of minicomputers of approximately the same level of capability into a network of hierachical computers for the purpose of data processing.

This concept of placing low-cost computer power at various action points in an organization, and linking these points where necessary, is happenning in EPA and deserves serious management consideration. Nearly 3 years ago, during the early discussions of the Cincinnati Laboratory Automation project, this concept of linking various levels of computers was presented. The issues related to distributed processing are significant and sometimes emotional; the concept is contrary to centralization and for this reason is bound to cause some confusion at the management level.

EPA is a decentralized organization with more than two thirds of its employees, or 6,000 people, in autonomous field locations; therefore, planning for distributed processing is imperative. However, to ensure that distributed processing does not get out of hand, the decentralization of ADP operations should be selective with strong centralization and control.

Figure 1 highlights the respective functional responsibilities to be assumed by the Headquarters Management Information and Data Systems Division and by the users. These responsibilities have been assumed in the Cincinnati Laboratory Automation project and are applicable to the generalized use of distributed processing in EPA.

An interesting trend can be seen in Figure 2. As computer technology is evolving, the respective functions of ADP and user departments are being exchanged. What were previously functions of ADP departments are being user department functions, and previous functions of user departments will soon be the responsibility of ADP departments.

As technology is developing, the most important and expensive resource in minicomputer applications is personnel. Equipment costs will continue to decrease but personnel costs will continue to increase. We have reached the point where one can purchase a microcomputer from petty cash and carry it in a pocket; but in order to effectively use this equipment, a person knowledgeable in hardware/software, logic design, interfacing techniques, and real time assembly language programing is essential.

In conclusion, as stated before, computing power will soon be essentially free. Many of the functions currently performed on a large central computer will be performed locally or in Regional centers. EPA will have networks of computers with distributed intelligence consisting of modular central processors with modular programs being executed in many different subprocessors.

Terminals will be superintelligent and computer peripherals will contain microprocessors. The technology which this represents shows every promise of bringing about the benefits that computers have been expected to deliver since their acceptance more than 20 years ago.

The proliferation of distributed processing in EPA need not be a threat to the concept of centralization; it only places more responsibility on all involved to communicate and coordinate more effectively. An unresponsive organization cannot keep minicomputer applications from developing but it may lose the benefits to be achieved now and may prevent easy expansion or integration in the future.

Management Information and Data Systems Division (MIDSD)

- Authority and responsibility for establishing policy and long-range goals
- Planning and project selection
- Coordination of:
    - Systems development
    - Hardware selection
    - Programing language selection
- Audit of decentralized operations
- Operation of large-scale computing facilities

Users

- Full-time participation on project teams
- Control of the systems design effort
- Hardware/software maintenance and operation
- Employment of programer/analyst

Figure 1
Functional Responsibilities

Era	Technology/Control Factors	Functions of User Departments	Functions of ADP Departments
1950-60	• Computers are new tools • Lack of understanding of computers	• Part-time participation on design of ADP systems • Maintain data base	• Select and maintain hardware • Control the design effort • Full-time participation on project teams • Employ programer/analysts
1960-70	• Proliferation of technology • Project teams emerge	• Full-time participation on project teams	• Select and maintain hardware • Control the design effort • Full-time participation on project teams • Employ analysts • Employ programers • Maintain data base
1970-75	• MIS • Teleprocessing • Better understanding of computer usage	• Control of project team • Full-time participation on project teams	• Select and maintain hardware • Full-time participation on project teams • Employ analysts • Employ programers • Maintain data base
1975-85	• Distributed processing • Corporate data base • High-level language	• Control of project team • Full-time participation on project teams • Employ analysts • Limited programing	• Select and maintain hardware • Part-time participation on design • Employ programers • Maintain data base
1985-90	• High-level application software for minicomputers	• Select and maintain hardware • Control the design effort • Full-time participation on projec teams • Employ programer/analysts	• Part-time participation on design • Maintain data base

Figure 2
Exchange of Functions

UNIVAC 1110 UPGRADE

By M. Steinacher

A major hardware upgrade is planned for the Univac 1110 computer system at the National Computer Center. The scheduled new equipment will be delivered and installed during mid-December and ready for production use in early January.

The reasoning behind the upgrade and the anticipated benefits are examined in this paper. A brief review of the Univac's procurement history is also included as background.

The original hardware/software specifications, ultimately used to procure the Univac computer, were prepared as early as September 1970. EPA had not yet been established and the procurement was intended to support the growing ADP requirements of the National Air Pollution Control Administration. As EPA and the RTP NERC were formed, the developing specifications were hastily modified to address the relatively unknown requirements of the new Agency and its local RTP components.

EPA was relatively inexperienced in the procurement of large-scale computer systems when, in early 1971, it engaged the General Services Administration (GSA) as an administrative and contracting agent in the procurement. EPA personnel maintained primacy as technical advisors and handled the technical aspect of the benchmarking and proposal evaluations.

GSA was particularly concerned about maintaining competitive procurement, and, in some instances, it was necessary for EPA to modify technical specifications to facilitate open bidding. Although it may have been very difficult for the Agency at that early date to absolutely justify certain capacity and speed requirements, GSA modifications may be directly responsible for the throughput bottlenecks experienced after installation.

However, open bidding competition was achieved, and three vendors were benchmarked in the fall of 1973. Univac passed all of the mandatory requirements and was by far the lowest bidder. It was awarded the contract in June 1973. The original configuration was subsequently installed in October 1973, and was finally accepted by the Agency in February 1974.

During EPA's early developmental years and while the procurement process was being executed, the ADP workload to be supported by the RTP computer grew at an accelerated pace. The expanding workload was further compounded by the developing interest in time-sharing applications. As a result, the actual computer requirements to be supported by the new Univac system were materially different from those reflected in the 1970/71 equipment specifications.

The initial performance of the Univac 1110 was certainly less than desirable. It was characterized by frequent stops, periods of degraded operation, and limited responsiveness. Although still less than stable, the system did settle down to reasonable periods of at least predictable, if not normal, performance. Subsequently, it was possible, through the use of hardware/software monitors, to evaluate and quantify performance.

Performance analysis clearly demonstrated that even under the most stable operating conditions, a serious hardware imbalance existed. Less than 30 percent of the available machine cycles were actually being used. The computer was literally waiting for work, work that was apparently bottlenecked somewhere. Individual jobs were in constant competition for the same peripheral device and for an adequate memory resource for program execution. As the operating system attempted to handle the situation by swapping jobs in and out, a significant overhead workload of at least 23 percent was being generated.

Recent monthly averages of 32 stops and 13.8 hours between failures indicated that the system was approaching operating stability. Assuming it could be achieved, a more productive hardware mix had to be considered. Therefore, negotiations with Univac addressed the inadequacies of mass storage and primary/extended memory as well as the need for dual access channels (I/O paths) and additional tape units.

The December upgrade reflects the response to these requirements and includes the following component increases:

.	Mass storage	68%
.	Primary storage	33%

- Extended storage 200%

- 8 additional magnetic tape drives

In addition, a second operator console and a second I/O control unit were added to provide for backup and for operating redundancy.

The future outlook for the Univac's computing potential appears bright. Univac estimates that the upgrade will significantly improve throughput by 50 percent and turnaround by 80 percent. In addition, at least 30 percent more jobs can be active and a similar percentage-increase in the number of demand terminals supported can also be realized. The unrealistic and unproductive overhead workload will be greatly reduced, thereby freeing even more computing potential.

We all look forward to the arrival of the Univac 1110 as a dependable and powerful Agency resource.

A CASE FOR MIDICOMPUTER-BASED COMPUTER FACILITIES

By D. Cline

INTRODUCTION

When the general purpose digital computer became a reality in the mid-1950's, only the largest corporate organizations could afford the capital outlay required for acquisition of a large-scale computer facility. The huge cost of acquiring and operating large-scale computer facilities was the factor that most delayed the development of small-scale computer facilities. Drastic cost reductions in hardware were realized in the early 1970's as a result of large scale integration (LSI) technology. Small-scale computer facilities became a reality with the introduction of the midicomputer which has an operating system that supports a multiprograming environment and has device-independent input/output.

LARGE-SCALE CENTRAL FACILITIES

The concept of economy-of-scale is the most powerful argument favoring large-scale central computer facilities. According to this concept, many users have collective access to a degree of computational power, where none would have access if the total cost of a facility had to be borne by each of the users. Other high cost, low usage, special peripheral devices may also be shared by numerous users in a large-scale computer facility.

In EPA, a common point of debate is accessing national data bases. Users of a large-scale central computer facility can access national data bases. Data bases often mentioned include water quality, air quality, library, financial, and personnel data. Another attractive feature of large-scale facilities is their large memory capacity, which is useful when executing large tasks. The remoteness of a user from the computer facility is usually of little concern as most large facilities may be accessed via direct dialing.

SMALL-SCALE LOCAL FACILITIES

On the other hand, many minicomputers and midicomputers of the mid-1970's provide a greater computing capability than was available from the largest computers of the mid-1950's. In fact, a complete small-scale system could be purchased for a year's budget that would support a large-scale time-sharing facility. In EPA, at least three feasibility studies performed to date substantiate this statement.

Although access to national data bases is an attractive feature, the requirement for utilizing national data bases in EPA's research laboratories varies from no use at all to daily use, depending on lab needs. Usually only portions of these data bases are required by any individual laboratory, in which case a subset of large data bases could be implemented on a local facility where intensive utilization is indicated.

Although many large-scale computer facilities have large memory capacities, at EPA's General Time Sharing Utility (GTSU) large memory segments are only readily available to the user overnight. Many computer programs that have a large memory requirement and an extremely small execution time are penalized by being forced to accept overnight response. This situation is normally acceptable for production jobs, but quite unacceptable for program development. Large computer programs can be executed on small machines by using external files for data storage and by using overlays to decrease the amount of memory required at any one time.

An attractive feature of EPA's GTSU is its extensive telecommunications network. To date, however, only 30 character per second (cps) lines are commonly available. On local facilities, the transmission rates are normally 30 to 1,250 cps.

In a research atmosphere where mathematical modeling and program development have high priorities, denial of access to the system after midnight on weekdays and all day Sunday severely impedes attainment of goals in a timely manner. With a local, in-house, small-scale facility, the system is available 24 hours a day, 7 days a week.

SUMMARY

Large-scale computer facilities have been the mainstay of the computer industry since their inception and will continue to provide for the bulk of computational requirements, mainly because of their favorable economics. There always will be requirements for maintaining large repositories of information, for executing huge memory-bound tasks, and for providing service for small users. The small-scale computer facilities, however, will complement the large-scale facilities in a cost-effective manner by performing much of the preediting of data before it is transferred to the large data bases, by executing many small routine tasks on a day-to-day basis, and by providing a means to conduct program development interactively. The net effect of small facilities will be to relieve the stress on large facilities. They will eliminate many of the small, routine tasks the large facilities normally encounter, and will enhance the ADP capability of the small research laboratory.

STATUS OF THE INTERIM DATA CENTER

By K. Byram

The EPA is currently in the midst of data center procurement. The contract awarded will furnish one of two data centers to be used nationwide by EPA.

When EPA was formed in December 1970, it incorporated several parts of predecessor agencies, and those agencies were receiving computer service from a variety of sources. For example, a computer center at Research Triangle Park (RTP), North Carolina, employed an IBM 360/50 to support EPA elements there. National Institutes of Health Computer Center in Bethesda, Maryland, and Boeing Computer Services in McLean, Virginia, were supporting Agency headquarters elements in the Washington, D.C. area and, to some extent, nationwide elements with IBM 370 series machines. Department of the Interior, Health Sciences and Mental Health Administration, Food and Drug Administration, Department of Agriculture, Atomic Energy Commission, and several universities were supplying computer services as well.

Two procurements were initiated in 1971 to 1972 to replace or upgrade the two principal computing resources listed above. The machine at RTP was to be replaced by a much larger system, and the contract at Boeing Computer Services was to be reopened to competition. Resulting awards were for a Univac 1110 at RTP, in February 1974, and a contract with Optimum Systems Inc. to replace the Boeing contract, in April 1973.

At about this time, General Electric (GE), under contract to EPA, completed a study which recommended that the headquarters workloads, then concentrated at OSI, NIH, and a few other places, be projected for transfer to an Agency-operated facility called the Washington Computer Center (WCC). GE also recognized a continuing need for a data center at Research Triangle Park to service EPA elements located there. EPA then began to plan a Washington Computer Center of its own to follow the OSI contract which would expire in 1975. An in-house task force was formed to coordinate the planning for the center. As the first step, the task force rejected the GE study's workload projections and basic conclusions, and began in 1974 to revalidate the study. Concurrently, the workload at NIH was transferred to OSI.

In the summer of 1974, recognizing the long leadtime for procurement of the large WCC data center, and recognizing that the OSI contract would expire soon, the Agency decided to reopen the OSI contract to competition. All of the studies to date had been oriented toward data center hardware; now, EPA's appropriations committee, investigating the Agency's growing ADP expenditure levels, demanded a 5-year plan including not only data centers, but information systems and staffing blueprints as well. Index Systems, Inc., was awarded the contract to perform this study. It recommended that the Agency continue, over the 5-year period, to operate the two data centers at Washington and RTP. They also suggested that the Agency exercise its purchase option on the Univac 1110 at RTP.

This study confirmed that reopening the OSI data center to competition was in order and should proceed. During the term of this new interim contract, EPA could complete and update its requirements studies, and procure a "permanent" data center, called the Washington Computer Center. Ever since EPA began, the workload which is now on the OSI data center had been satisfied with IBM 360/370 series equipment. To avoid massive conversion problems, EPA wished to specify IBM equipment for the interim center.

However, GSA was charged with implementing a full competition policy. This policy philosophy states that it is in the interest of the Government to have several computer manufacturers providing hardware in a competitive environment. Continued reliance for decades on one vendor by agencies citing difficulties of conversion to other vendors' equipment could place the Government at great disadvantage in receiving price and service from that vendor. Yet GSA's authority, contained in the Brooks Bill, prevents them from interfering with an agency's mission. Since requiring full competition and a probable conversion would interfere, GSA has found it difficult to prevent brand name specification in data center procurements. As a compromise, GSA has allowed "interim procurements" specifying brand names, to be followed by fully competitive procurements within 2 years, or longer if GSA and the agency mutually agree. In essence, GSA recognized the conversion problem. Instead of requiring that EPA encounter it with every procurement, GSA allowed EPA to have an interim period of brand name specification to develop and complete a conversion plan.

EPA specified IBM brand name equipment, and its historical situation mandated that the procurement serve EPA nationwide, with a majority of workload at headquarters, and almost none of it at RTP. Two aspects of this procurement set it apart from others. First, it is in three separately awardable parts. Second, it is a facilities management arrangement, for which the vendor (or vendors) is reimbursed for the cost plus an award fee.

The RFP's three parts are the data center hardware/software, the communications network, and the user support service. Each part is separately awardable, although one vendor could propose and win in all parts. While coordination could be a problem with separate vendors, best support in each area is not necessarily available from a single vendor. Also, one vendor could tend to emphasize his management strength in the area of most profit, with adverse effect on the other areas. One interesting aspect of the communications network portion of the RFP is that it specifies a network to provide nationwide access to the WCC, as well as to the Univac data center at RTP.

The services are to be procured under a facilities management concept. In essence, EPA and the vendor (or vendors) will jointly determine the equipment, network components, and level of staffing necessary for the three parts of the center. The vendor will then obtain the necessary resources and operate them to EPA specification. This differs from the current pricing arrangement for the OSI contract, whereby EPA is billed essentially job-by-job for the work accomplished.

The procurement was forwarded for GSA approval in December 1974 and approved in July 1975. The procurement was released to the public in August, with proposals due on October 31. Evaluation is proceeding, and an award is expected by September 1976.

LARGE SYSTEMS VERSUS SMALL SYSTEMS

By R. W. Andrew

The primary consideration in planning development of ADP resources for the U.S. Environmental Protection Agency (EPA) should be the needs of the ultimate user community, that is, the EPA scientists and administrators. The following discussion, representing the viewpoint of a Research and Development (R&D) scientist-user, is intended to describe some of those needs and how they can best be met by future developments in ADP resources for EPA. Until recently, such resources in ADP "just growed," like the proverbial Topsy. This opportunity to present and discuss our needs heralds a welcome change in management philosophy toward fitting the resources to the job rather than the reverse.

Perhaps the best way to begin to outline future needs is to describe those present needs which are unmet. At present, ADP resources for EPA are supplied largely by the maxicomputer systems at Optimum Systems Inc. (OSI) and Research Triangle Park (RTP). Both systems are designed and operated primarily for large-scale batch jobs and large data bases such as STORET and AEROS. Remote time-share users, although now garnering an increasing proportion of total usage, have received relatively little attention in the planning and operation of the systems. Remote user access to scientific and statistical software packages has been added as an afterthought, with little or no previous planning or overall design constraints. The remote time-share user has been forced to operate in a "batch-mode" environment, or to pay an excessive premium for time-sharing operation (TSO) or demand operation which accesses only part of the software packages. These are the precise services and software, however, most needed by the remote scientist-part-time programer. Perhaps the best example illustrating this point is that none of the present EPA-supported computer services offer online time-share compilers or interpreters for BASIC language programs, yet BASIC is rapidly becoming the unviersal language of the scientist-programer.

A second need, perhaps the major stumbling block to the R&D scientist-user of the present systems, is learning the terminal operation and Job Control Languages (JCL). The typical scientist has neither the inclination nor the time required to learn an additional language or the complex JCL required to run his programs. While the present EPA systems provide raw computer power equal to any conceivable task, these systems are virtually unapproachable by the scientist-user, laboratory, or office lacking trained computer personnel.

What ought to be done to satisfy these and future needs of the R&D scientist-user community? First, it is important to expose economy-of-scale for the myth that it is. Because of the rapidly declining cost of large-scale memories and microprocessors, the economies of central processors are rapidly disappearing. Such economies are readily outweighed by the high cost of communications, project delays, and specialized training. As a result, most EPA laboratories are following the lead of their industrial counterparts and are turning to the use of dedicated minicomputers. In some cases the minicomputers have more flexibility and dedicated memory than is available through the large systems. This transition is happening so rapidly that it precludes the best prior planning and management efforts. An estimated 75 to 100 minicomputer systems are presently installed in EPA laboratories and offices; most without any centrally coordinated planning, design, or sanction. These systems, however, do satisfy, though somewhat inefficiently, the need for rapid, cost-effective computation. Therefore, present planning should recognize and encourage the use of such decentralized computer facilities, and should provide for some standardization of equipment, software, and training as a means of improving the operational efficiency within R&D. An additional means of improving both flexibility and efficiency would be planning and instituting distributive networks of such minisystems, combined with the larger Agency-wide systems.

Another possible means of providing improved computer service to all of the EPA user community would be to institute greater segregation and dedication of operating system software (and possibly hardware) to specific tasks and levels of computation. For example, it is inefficient to require a scientific user running a small FORTRAN job to utilize the same terminal and JCL as the STORET user wishing to manipulate and sort massive data bases. The FORTRAN user should be able to edit, compile, execute, and save his job with simple commands such as: FORT, LIST, RUN, and SAVE, and all machine transactions and record keeping should be invisible to the user. This type of operation is an established fact with most university or science-oriented computer systems.

Some additional suggestions to help satisfy existing and contemplated needs are:

- Increased use of "smart" terminals; e.g., TEXTRONIX 4051, H-P 9830, or IBM 5100 as interpreters; or preprocessors for on-line data reduction and formatting from active experiments
- Formation of scientific and/or statistical software user groups for formal sharing of programs and problem solutions
- Institution of computer aided instruction (CAID) for the training of novice users
- A switch from contractor-supported to EPA-supported user service at the central agency computer (OSI).

While it is obvious that EPA's need for and usage of ADP will continue to increase at a geometric rate, we must recognize that those needs and usages are becoming more and more specialized. Consequently, the ability of a centralized computer utility to be all things to all users becomes increasingly difficult. In the future, planning should avoid equating size with flexibility, and specialization should be anticipated in such a way that it can be a benefit to EPA.

SUMMARY OF DISCUSSION PERIOD - PANEL VI

The questions after the session, Future Developments in ADP Resources for EPA, focused on six primary topics. They are discussed below.

Policy

Questions addressed general Agency policy and, more specifically, ADP policy as established by the Management Information and Data Systems Division (MIDSD). One query raised concerned which level ADP issues should address. The Assistant Administrator (AA's) are involved in long-range planning of ADP requirements and resources. An ADP steering committee comprised of the AA's is currently developing a 5-year ADP plan for the Agency. The Deputy Assistant Administrator (DAA's) are involved in budget decisions regarding ADP, as are Laboratory Directors. ADP Coordinators within each of the laboratories are responsible for communicating information, both administrative and technical, within their respective laboratories and for interacting with, and receiving guidance from, the Office of Research and Development (ORD) ADP Coordinator.

Each ORD Laboratory has been encouraged to develop a plan for the future, of which ADP certainly can be a part. This plan may show a strong in-house effort. In this case, there would be a lower travel budget, lower grade-point average, more technicians, and less PhD's. This is the opposite of a strong out-of-house program.

Concern was expressed regarding the possibility of MIDSD's deciding who could establish stand-alone computer capabilities. A policy has not been generated. In any event, the decision would depend upon the users' needs.

Concern was also expressed regarding strong MIDSD alignment with the General Services Administration (GSA). MIDSD has made a major effort to follow GSA policy, but EPA is allowed by GSA to do what it feels necessary.

Approval by MIDSD is required for procurement of items under Federal Schedule 66, Laboratory Instruments and Automation.

National Computer Center

The first issue concerning the National Computer Center (NCC) located at Research Triangle Park, North Carolina, was the architecture of the Univac 1110 as it relates to the present mix of demand use and batch use. This mix is user specific as opposed to data center specific. The data center can, however, limit what users do over demand as opposed to what they do over batch.

The specifications for the Univac 1110 include support for both demand use and batch use. A batch processor was not purchased. A larger amount of slow memory is being acquired to handle demand use more effectively. Pricing is being altered to make batch use cheaper than demand use.

The types of use for which demand access can better be used include the retrieval of information from data banks and computational functions which are time-specific and dependent. Program developing may be done more effectively on minicomputers.

The role of NCC managers in the selection of the interim and a permanent computer facility was questioned. One of the NCC staff is on the interim center evaluation panel. Future determinations concerning a permanent center, which is not geographically or machine restricted, are open for participation.

The NCC will become more involved than before with the remote users. Each regional office is being provided with 2-day seminars on the use of systems on the Univac 1110. However, users at RTP will receive service at the same, or an increased, level as compared to past service.

Contracting

There has been a shift to using contractor personnel to operate Agency facilities. Concern was expressed regarding the potential loss of Federal skills to contractor employees. The Agency will continue the functions of management, including the technical planning of resources. Contractor personnel will be used to implement specific projects or to provide their expertise for a product containing factual data. The current large budgets will encourage a tendency to use onsite contracting. There probably will be no new positions, and if jobs are facility operations in nature, contractors may be used.

Communications

A national communications system will extend to all Regional Offices and to the laboratories located at RTP, Cincinnati, Las Vegas, and Corvallis. The network is designed to handle communications to both Optimum Systems Incorporated (OSI) and the National Computer Center.

A user's catalog may be generated which is accessible to anyone. This is contingent upon standardization of programs to run on any EPA machines. This would allow, in simple terms, the use of any particular machine to solve a problem rather than the use of a wide mix of machines.

System Size

There was great interest in the use of large computer systems versus small systems. At this time EPA does not have a firm policy regarding minicomputers. There is no deliberate intent to diminish the number of minicomputers. However, it will be ascertained that available purchased capacity of a system is used before additional capacity is purchased.

There was evident support for a PDP 10 system, which is much more approachable by a scientist. The suggestion was well received that ORD should buy a PDP 10 for use by scientists.

A mix of different computers operated by EPA was addressed as being desirable. EPA could then borrow programs without conversion problems. This mix could include the following: IBM 370, CDC 6000, Univac 1110, and PDP 10.

A smaller system application used for reducing and feeding data into a higher level system for large summaries may be efficient. These summaries may include trends, predictions, and historical displays. This would involve a stand-alone computer with concurrent terminal capabilities for interaction with a larger system; in other words, distributive processing.

Distributive Processing

Support was enlisted for the distributive processing concept, inherent in the pending standard terminal procurement. This procurement specifies a standard terminal consisting of a minicomputer with stand-alone capabilities and concurrent terminal capabilities.

This would be a small machine that could hook up with OSI, or NCC, or other centers. Data then could be easily provided to central data banks, and programs on a distant central system could be used locally.

Pooling the power of individual minicomputers with the large systems would result in a positive synergism. One problem is increased local processing "creep." Through distributive processing, a local operator will learn more about the software capability available from the larger system, increase the size of the local system to maintain local control over the acquired software, and slowly generate a maxisystem locally which then becomes independent of the distributive system.

Implementing distributive processing within the Agency, because of its size, would be a formidable task. However, individual components such as research laboratories could begin. This could be done, first, by defining functions being performed by the research program at the particular locations; second, by identifying the data processing needs; and third, by analyzing the data acquisition process, the analytical or data reduction processes that must be executed, and the level at which they should be performed. The use of second and third shift time available on larger systems should not be overlooked. With less demand use, it is more stable. With this information available, local systems could be defined in relation to the use of larger systems.

APPENDIX

LIST OF ATTENDEES

Allison, G.
Environmental Monitoring and Support Laboratory
Environmental Protection Agency
P.O. Box 15027
Las Vegas, Nevada 89114

Almich, B.
Computer Services and Systems Division
Office of Administration
Environmental Protection Agency
Cincinnati, Ohio 45268

Anderson, G.
Health Effects Research Laboratory
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Andrew, R.
Environmental Research Laboratory
Environmental Protection Agency
6201 Congdon Boulevard
Duluth, Minnesota 55804

Barton, G.
General Chemistry Division, L-404
Lawrence Livermore Laboratory
P.O. Box 808
Livermore, California 94550

Berger, J.
Department of Commerce
National Oceanic and Atmospheric Administration
Environmental Data Service
Washington, D.C. 20235

Borthwick, P.
Environmental Research Laboratory
Environmental Protection Agency
Gulf Breeze, Florida 32561

Bryan, S.
Health Effects Research Laboratory
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Budde, W.
Environmental Monitoring and Support Laboratory
Environmental Protection Agency
Cincinnati, Ohio 45268

Byram, K.
Management Information and Data Systems Division (PM-218)
Office of Planning and Management
Environmental Protection Agency
Washington, D.C. 20460

Chamblee, J.
ADP Coordinator
Office of Water Program Operations (WH-547)
Office of Water and Hazardous Materials
Environmental Protection Agency
Washington, D.C. 20460

Cirelli, D.
Ecological Monitoring Branch
Technical Services Division (WH-569)
Office of Pesticide Programs
Environmental Protection Agency
Washington, D.C. 20460

Cline, D.
Environmental Research Laboratory
Environmental Protection Agency
College Station Road
Athens, Georgia 30601

Conger, C.
Monitoring and Data Support Division (WH-553)
Office of Water and Hazardous Materials
Environmental Protection Agency
Washington, D.C. 20460

Couch, J.
Environmental Research Laboratory
Environmental Protection Agency
Sabine Island
Gulf Breeze, Florida 32561

Davies, T.
Environmental Research Laboratory
Environmental Protection Agency
Sabine Island
Gulf Breeze, Florida 32561

Dell, R.
Central Regional Laboratory
Environmental Protection Agency
Region V
1819 West Pershing Road
Chicago, Illinois 60609

Enrione, R.
Health Effects Research Laboratory
Environmental Protection Agency
Cincinnati, Ohio 45268

Fairless, W.
Central Regional Laboratory
Environmental Protection Agency
Region V
1819 West Pershing Road
Chicago, Illinois 60609

Frazer, J.
Department of Electrical Engineering
Colorado State University
Fort Collins, Colorado 80521

Goldberg, N.
Environmental Research Laboratory
Environmental Protection Agency
South Ferry Road
Narragansett, Rhode Island 02882

Greaves, J.
Department of Electrical Engineering
Southeastern Massachusetts University
North Dartmouth, Massachusetts 02747

Hammerle, J.
Monitoring and Data Analysis Division
Office of Air and Waste Management
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Hart, J.
Computer Services and Systems Division
Office of Administration
Environmental Protection Agency
Cincinnati, Ohio 45268

Hertz, M.
Health Effects Research Laboratory
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Johnson, M.
National Computer Center (MD-34)
Environmental Research Center
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Jurgens, R.
Environmental Sciences Research Laboratory
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Kelsey, A.
Health Effects Research Laboratory
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Kinnison, R.
Environmental Monitoring and Support Laboratory
Environmental Protection Agency
P.O. Box 15027
Las Vegas, Nevada 89114

Kleopfer, R.
Environmental Protection Agency
Region VII
1735 Baltimore Street
Kansas City, Missouri 64108

Knight, J.
Health Effects Research Laboratory
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Kolojeski, P.
Office of Air, Land and Water Use (RD-682)
Office of Research and Development
Environmental Protection Agency
Washington, D.C. 20460

Kopfler, F.
Health Effects Research Laboratory
Environmental Protection Agency
Cincinnati, Ohio 45268

Koutsandreas, J.
Office of Monitoring and Technical Support (RD-680)
Office of Research and Development
Environmental Protection Agency
Washington, D.C. 20460

Krawczyk, D.
Environmental Research Laboratory
Environmental Protection Agency
200 S.W. 35th Street
Corvallis, Oregon 97330

Lawless, T.
Environmental Monitoring and Support Laboratory
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Lawrence, C.
Office of Monitoring and Technical Support (RD-680)
Office of Research and Development
Environmental Protection Agency
Washington, D.C. 20460

Lowrimore, G.
Health Effects Research Laboratory
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Meyer, R.
International Research and Technology
1501 Wilson Blvd.
Arlington, Virginia 22209

Mullin, M.
Grosse Ile Laboratory
9311 Groh Road
Grosse Ile, Michigan 48138

Myers, M.
Office of Research and Development (RD-672)
Environmental Protection Agency
Washington, D.C. 20460

Nime, E.
Computer Services and Systems Division
Office of Administration
Environmental Protection Agency
Cincinnati, Ohio 45268

Ott, W.
Office of Monitoring and Technical Support (RD-680)
Office of Research and Development
Environmental Protection Agency
Washington, D.C. 20460

Rhodes, R.
Environmental Monitoring and Support Laboratory
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Richards, N.
Environmental Research Laboratory
Environmental Protection Agency
Sabine Island
Gulf Breeze, Florida 32561

Risley, C.
R&D Representative
Environmental Protection Agency
Region V
230 South Dearborn Street
Chicago, Illinois 60604

Schoor, P.
Environmental Research Laboratory
Environmental Protection Agency
Sabine Island
Gulf Breeze, Florida 32561

Ustaszewski, Z.
Office of Health and Ecological Effects (RD-683)
Office of Research and Development
Environmental Protection Agency
Washington, D.C. 20460

Wheeler, V.
Environmental Monitoring and Support Laboratory
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Whitley, S.
Earth Resources Laboratory
National Aeronautics and Space Administration
Bay St. Louis, Mississippi 39520

Williams, R.
Municipal Environmental Research Laboratory
Environmental Protection Agency
Cincinnati, Ohio 45268

Worley, D.
National Computer Center (MD-34)
Environmental Research Center
Environmental Protection Agency
Research Triangle Park, North Carolina 27711